Self Correcting Tracking of Moving Objects in Video

7-19-2011

Niels Da Vitoria Lobo
University of Central Florida

Mehmet Caglar
University of Central Florida

Find similar works at: https://stars.library.ucf.edu/patents

University of Central Florida Libraries http://library.ucf.edu

Recommended Citation

https://stars.library.ucf.edu/patents/502

This Patent is brought to you for free and open access by the Technology Transfer at STARS. It has been accepted for inclusion in UCF Patents by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.
SELF CORRECTING TRACKING OF MOVING OBJECTS IN VIDEO

Inventors: Niels da Vitoria Lobo, Winter Park, FL (US); Mehmet B Caglar, Oviedo, FL (US)

Assignee: University of Central Florida Research Foundation, Inc., Orlando, FL (US)

Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1090 days.

Patent No.: US 7,983,448 B1
Date of Patent: Jul 19, 2011

Apparatus, methods, systems and devices for identifying and tracking a human hand in a video. The steps for identifying the hand include detecting parallel lines (bars), detecting curved fingertips (curves), grouping the detected parallel lines and curved fingertips according to a parallel line decision tree and a curved fingertip decision tree, respectively, and merging the parallel line group and the curved fingertip group to identify candidate hands.

19 Claims, 9 Drawing Sheets

ABSTRACT

Apparatus, methods, systems and devices for identifying and tracking a human hand in a video. The steps for identifying the hand include detecting parallel lines (bars), detecting curved fingertips (curves), grouping the detected parallel lines and curved fingertips according to a parallel line decision tree and a curved fingertip decision tree, respectively, and merging the parallel line group and the curved fingertip group to identify candidate hands.

19 Claims, 9 Drawing Sheets
OTHER PUBLICATIONS


* cited by examiner
START

DETECT PARALLEL LINES 110

GROUP PARALLEL LINES 130

DETECT FINGERTIP CURVES 120

GROUP FINGERTIP CURVES 133

MERGE GROUPS 140

TRACK HANDS 150

Fig. 2a
FROM STEP 110

CREATE N GROUPS OF PARALLEL LINE PRIMITIVES 131

DETERMINE PARALLEL LINE GROUP MEMBERSHIP 132

VOTE AND ELIMINATE GROUPS BASED ON VOTE 133

ELIMINATE GROUPS SUBSETS AND SMALL GROUPS 134

MERGE PARALLEL LINE AND FINGERTIP CURVE GROUPS 140

TO STEP 150

FROM STEP 120

CREATE M GROUPS OF FINGERTIP CURVE PRIMITIVES 136

DETERMINE FINGERTIP CURVE GROUP MEMBERSHIP 137

VOTE AND ELIMINATE GROUPS BASED ON VOTE 138

ELIMINATE GROUPS SUBSETS AND SMALL GROUPS 139

Fig. 2b
FROM STEP 140

INITIALIZE TRACKER 151

TRACK FORWARD AND BACK 152

RECORD TRAJECTORY ELEMENTS 153

MERGE AND ASSIGN WEIGHT TO MERGED TRAJECTORY ELEMENTS 154

GENERATE TRAJECTORY HISTOGRAM 155

EXTRACT AND SEPARATE HANDS 158

STEP 156

Fig. 2c
START

DIVIDE VIDEO INTO FRAMES 162

DETERMINE OBJECT VELOCITY 164

COMPARE OBJECT & HAND VELOCITY 166

NO

CHANGE ? 168

YES

MARK AS AN EXCHANGE 170

FIG. 2d.
Fig. 8

Actions

One Hand
Hand without object
Move
Pick up object
Put down object

Two hands interact
Hand with object
Hand Shaking
Object Exchange
Move
SELF CORRECTING TRACKING OF MOVING OBJECTS IN VIDEO

The application claims the benefit of priority to U.S. Provisional Patent Application No. 60/810,284 filed on Jun. 2, 2006.


FIELD OF THE INVENTION

This invention relates to video surveillance and, in particular, to methods, systems, apparatus and devices for identifying and tracking of moving objects including human hands in video for uses such as human computer interaction applications, surveillance, gesture recognition and understanding human-human interactions.

BACKGROUND AND PRIOR ART

Tracking in unconstrained and cluttered environments is a difficult task, especially when the object of interest is articulated. When a tracker fails, re-initialization is needed. In systems that are not self initialized, the initial structure from the manual initialization is assumed to be preserved over time for self correction. Even if an online model of the object is kept, it is not easy to determine automatically if the tracking has failed and when.

Tracking has been a well researched area in computer vision. There are numerous tracking systems such as Kanade-Lucas-Tomasi (KLT) Feature Tracker described in C. Tomasi and T. Kanade. Detection and tracking of point features, Technical Report CMU-CS-91-132, Carnegie Mellon University (April 1991) and mean shift approach described in D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis, Pattern Analysis and Machine Intelligence, 24(5):603-619, (2002). The common problems of the prior art algorithms is initialization and self-correction.

Another problem with prior art detectors is that they have a high number of false positives (approximately 2 false positives per detection) and detection is not possible in some frames due to self occlusion and inability to detect the object of interest, such as finger primitives in low resolution images. Hand tracking is important in many applications including HCI, surveillance, gesture recognition and understanding human-human interactions. Prior art systems fail to track hands in uncontrolled environments, limiting them to simple gesture recognition problems with uniform background. In these systems, initialization usually comes from stationary hands in the system either by observing the hand for a short time frame, building pre-computed skin color histogram by training or manual initialization. Most prior art systems are unable to recover when the tracking fails.


The skin detector is either trained on the skin color that is obtained from the face R. Rosales, V. Athitsos, L. Sigal, and S. Sclaroff. 3d hand pose reconstruction using specialized mappings, In Proc. Intl. Conf. Computer Vision, (2001), pp. 378-387, or as in the more frequent case, on skin color extracted from the training images as described in Bretzner et al, Hand gesture recognition using multi-scale color features, hierarchical models and particle filtering, In Proc. Intl. Conf. Hand and Gesture Recognition, pp. 423-428, (2002). However, the prior art systems fail under poor lighting conditions or when tracking skin colors that they are not trained for.


Kolsch and Turk described a system for detecting hands based on Ada-Boost. Their approach is view-specific and limited to a few postures of the hand. Yuan disclosed a hand detector, where detection of the hand successfully depends on continuous movement of the hands, therefore requires video as input. Moreover, in this video the hand posture is required to change, such that there is no correspondence between the hand regions among the frames that are being compared. Bretzner et al. use a hand detector before recognizing the hand posture; however their system performs poorly if a prior on skin color is not used.


Moreover, the hand is assumed to extend from the right hand side of the image, and assumed to exist in the scene. They use “U” shaped fingertip patterns to detect fingertips in already detected hands. Another hand tracker to mention is disclosed by Rehg et. al., Visual tracking of high dof articulated structures: An application to human hand tracking, In Proceedings of the 3rd European Conference on Computer Vision (ECCV ’94), volume II, (May 1994), pp. 35-46 where the authors describe methods and systems that track simple features of the hand instead of tracking the hand in its entirety. However, they use a simple background for tracking.

SUMMARY OF THE INVENTION

A primary objective of the invention is to provide new methods, systems, apparatus and devices for identifying the location of open hands in video for use in tracking moving hands.

A secondary objective of the invention is to provide new methods, systems, apparatus and devices for identifying and tracking of moving objects including human hands in video using primitives corresponding to the object being tracked.
A third objective of the invention is to provide new methods, systems, apparatus and devices that employing finger-tip curve finders and detectors of tubular shapes that resemble fingers to locate open hands in video.

A fourth objective of the invention is to provide new methods, systems, apparatus and devices that use a competition strategy to propagate fingertip identification information across frames of a video.

A fifth objective of the invention is to provide new methods, systems, apparatus and devices to provide simple hand tracking without the need for training for skin color.

A sixth objective of the invention is to provide new methods, systems, apparatus and devices that eliminates the false positives while filling in the missed detections by calculating the most probable track the hand is likely to have followed.

A seventh objective of the present invention is to provide new methods, systems, apparatus and devices for a self initializing and self correcting tracking technique that is robust to different skin color, illumination and shadow irregularities.

An eighth objective of the present invention is to provide new methods, systems, apparatus and devices for tracking that is well suited to uncontrolled environments for successfully tracking detected objects in the sequence where detection does not necessarily have to be successful in each frame observed.

A ninth objective of the present invention is to provide new methods, systems, apparatus and devices for tracking hands by first finding the finger primitives in the image then combines them to find possible hands. This enables it to be more generic and not dependent on skin color. Then tracking is done by dividing the motion into two components as local and global motion and recomputing the hand features at each frame.

A first embodiment provides a method for detection and tracking of a moving object including the steps of detecting at least two primitive classes for identifying the moving object within plural frames of video to generate plural primitive candidates for each class, grouping each detected primitive candidates in each class separately according to a decision tree corresponding to each primitive class, merging primitive groups for each class of primitives to identify the object being tracked, and tracking the object through the plural frames according to the merged primitive groups.

The detection step includes detecting plural primitives candidates for one of the primitive classes, extracting a parameter of a model of the one of the primitive classes, superimposing the primitive model parameter on each primitive candidate of the class and comparing each primitive candidate to the primitive model to determine which primitive candidate is a primitive.

Once the primitive candidates have been detected, they are grouped. The grouping step includes creating one group for each detected primitive candidate in each class to generate plural groups of primitives candidates for each primitive class using the decision tree corresponding to the primitive class, determining group membership of each of the plural groups from the plural primitive candidates detected in the detecting step, voting to eliminate primitive groups based on voting results and eliminating subset groups and small groups to reduce false primitive positives. The grouping step yields primitive candidates locations for each primitive class. In a next step, the primitive groups of a first class are merged with primitive groups of the second class and it is determined which merged groups include the object. For the groups having the detected object, the object is tracked through the plural frames by tracking the object forward and backward through each of plural sets of video frames to obtain plural trajectories of the object, recording each of the plural trajectories, merging the plural trajectories, assigning weights to each of the merged trajectories, generating a trajectory histogram from the merged trajectories and lastly, extracting and separating the hands from the trajectory histogram.

A second embodiment of the invention provides a method for detection and tracking of moving hands in a video. First parallel line primitives and curved fingertip primitives are detected then the parallel line primitives are grouped according to a line decision tree and curved fingertip primitive according to a fingertip decision tree. The parallel line groups and curved fingertip groups are merged to identify the hand and the hand is tracked through the plural frames according to the merged primitive groups.

The hands are tracked through a series of video frames by tracking the object forward and backward through each of plural sets of video frames to obtain plural trajectories of the object, recording each of the plural trajectories, merging the plural trajectories, assigning weights to each of the merged trajectories, generating a trajectory histogram from the merged trajectories and extracting and separating the hands from the trajectory histogram.

Further objects and advantages of this invention will be apparent from the following detailed description of preferred embodiments which are illustrated in the accompanying drawings.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Before explaining the disclosed embodiments of the present invention in detail it is to be understood that the invention is not limited in its application to the details of the particular arrangements shown since the invention is capable of other embodiments. Also, the terminology used herein is for the purpose of description and not of limitation.

The following is a list of designators used throughout the description to identify elements shown in the drawings.

- hand
- parallel lines
- fingertip curves
- background
The apparatus, methods, systems and devices of the present invention solve the prior art initialization and self-correction problems with an approach that exploits the fact that video is available in its entirety and there exists a weak detector that finds an object of interest in a number of frames. For the purpose of illustration, and not limitation, the novel technique disclosed in this application is applied to hand tracking. Hands are articulated objects with high degree of freedom, and they are hard to detect in cluttered backgrounds such as the background shown in FIG. 1 and FIG. 6.

Hand detection is a preliminary step to a number of applications including HCl, surveillance, gesture recognition, hand tracking and understanding human-human interactions. In regard to hand tracking, the present invention solves the false-positive while filling in missed detections by calculating the most probable track the hand is likely to have followed. The system finds the globally optimum track for the hands by looking at the video in its entirety. If the detection of the hand is successfully carried out in some of the frames in the video, then the missing pieces are completed according to the tracking. In other words, the system finds the hands individually in each frame and the output of the tracking is used to fill in when the detection fails. The tracking method and system of the present invention is well suited to uncontrolled environments to successfully track detected objects in the sequence where detection does not necessarily have to be successful in each frame observed. False detections are also handled automatically, as long as they do not persist over time as is the general case.

The apparatus, methods and systems provide a self-initializing and self-correcting tracking technique that is robust to different skin color, illumination and shadow irregularities. Self initialization is achieved from a detector that has relatively high false positive rate. The detected hands are then tracked backwards and forward in time using mean shift trackers initialized at each hand to find the candidate tracks for possible objects in the test sequence. Observed tracks are merged and weighed to find the real trajectories. Simple actions are inferred extracting each object from the scene and interpreting their locations within each frame. Extraction is possible using the color histograms of the objects built during the detection phase. Applying the technique according the present invention to simple hand tracking had good results, without the need for training for skin color.

Hand detection is an essential component in initialization and self correction. The method, system, apparatus and device of the present invention provides an algorithm that uses finger primitives to detect hands in each frame. Currently there are two primitive classes of interest: parallel lines and fingertip curves. Alternatively, additional primitive classes and corresponding decision trees can be used as long as a sound algorithm to merge the results is determined.

Detection of Curved Fingertips (Curves):
The curved fingertip detection step 120 uses a variant of the person finding algorithm described in Garcia et al., Finding heads in single images, In Proc. Canadian Conf. Computer and Robot Vision, (2005) for finding the curved fingertips. This algorithm is a cropped version of prior art model described in T. Zhuo and R. Nevatia, Stochastic human segmentation from a static camera, In Proc. IEEE work, Motion and Video Computing, pp. 9 14, (2002). As the first step in detecting curves, the model edges are extracted from each model. Neighboring pixels of the same orientation are grouped together and the middle point of each group is marked as a fingertip top candidate.

The next step is superimposing the models on each candidate in the test image. This process yields how good the match is between the model and the test area. If an ideal fingertip is superimposed on a fingertip candidate 220, the k pixels of this ideal fingertip’s contour has coordinates m1, ..., mk and unit normals ν1, ..., νk. The score for a finger candidate is then given by the formula:

\[ S(x,y) = \sum_{i=1}^{k} a \cdot d_i(w_i(x,y)) \]

where D(x,y) is the distance to the closest edge pixel from (x, y), C(x, y) is the closest edge pixel from (x, y) and O(X, y) is the unit gradient vector at the pixel (x, y).

In order to maintain the scale invariance, the models are scaled along each dimension, and every scaled model is matched to the candidate. If a model at a specific threshold scores more than other scales and surpasses a given threshold, then the corresponding candidate on which the model is superimposed is marked as a fingertip. The hand images in FIG. 4 show an example of an output of step 120.

Steps 130 and 135: Grouping Primitives

After finding finger parallel lines primitives and curved fingertip primitives, the parallel line 210 and curved fingertip 220 primitives are grouped together in steps 130 and 135 respectively, using decision trees trained separately for each primitive class, parallel bars and curved fingertips, based on the relative properties of these primitives. Relative properties are those properties defined in such a way that different responses at different scales are constant. This avoids using individual properties of primitives. To maintain the scale variance, the ratio of the distance between two parallel lines to the width of the thinner line or the ratio of the radius of the thinner fingertip to that of the wider fingertip can be used. These are example features that are invariant for different scales and used for building the decision trees.

The question that is propagated through the decision tree is, given two detected primitives, parallel lines and curved fingertips in this example, whether they can belong to the same hand or not. Due to the nature of decision trees, the answer to this question is binary 1/0 (yes/no). According to the present invention, the system employs two decision trees, one decision tree trained for each of the finger primitive classes, parallel lines and fingertip curves. Alternatively, additional primitive classes and corresponding decision trees can be used as long as a sound algorithm to merge the results is determined.
Once it is possible to correctly identify whether or not it is possible for two given detected primitives to belong to one hand, groups of primitives are formed. The grouping process for each finger primitive creates groups of primitives employing the specific tree for that primitive class. The parallel line grouping and fingertip curve grouping steps 130 and 135, respectively, are shown in FIG. 2a. As shown, in step 131 the detected parallel line primitives 210 from step 110 are grouped and in step 136 the detected fingertip curve 220 primitives from step 120 are grouped. If there are N bar primitives detected, N groups are formed in step 131 for the detected parallel line primitives. If there are M curve primitives detected, M groups are formed in step 136 for the detected curve primitive. Each bar primitive and each curve primitive is the representative of a different group.

For each single group created in steps 131 and 136, the membership of a specific primitive is determined by the decision tree used in steps 132 and 137, a voting scheme is employed in steps 133 and 138 to eliminate primitives that do not get the majority vote within each group. Next, in steps 134 and 139, groups that have at least one supererset are eliminated and groups that have less than 3 primitives are discarded. An example output of this grouping steps is shown in FIG. 4. This representation is intermediate and shows the grouping of parallel lines 210 into hands 200.

Step 140: Merging Groups of Primitives

The grouping steps yield candidate hand locations for each primitive class. As shown in FIG. 6, each candidate hand is marked with a rectangle. First, assume that the p<sub>1</sub> group from finger primitive class p<sub>1</sub>, is called p<sub>1</sub>. In other words, the detected members of primitive class p<sub>1</sub> in the test image is p<sub>1</sub> through p<sub>n</sub>, where n is the number of groups that are found in steps 131 and 136, for primitive class p<sub>1</sub>, detected primitives from the grouping process will be p<sub>1</sub> through p<sub>m</sub>, where m is the number of groups that results from grouping, and so on. As many groupings of primitives as necessary can be denoted in this manner. The present invention uses two primitive classes, bars (parallel lines) (p<sub>1</sub>) and fingertip curves (p<sub>2</sub>). When comparing group f from p<sub>1</sub> to group g from p<sub>2</sub> to get a measure of how good a match f and g are, the present invention uses the scoring function:

\[
\text{score}(f,g)=\frac{\text{no Intersections}(f,g)^2}{\text{widthDiff}(f,g)^2+1}
\] (2)

where no Intersections and widthDiff f are as follows:

\[
\text{noIntersections}(f,g) = \sum_{r} \sum_{p} \text{intersects}(r, p)
\]

\[
\text{widthDiff}(f,g) = \sum_{r} \sum_{p} \text{intersects}(r, p)
\]

where intersects(r,p) is a binary function that returns 1 if the bar r intersects the fingertip curve p, and returns 0 otherwise.

Although a bar (set of parallel lines) can intersect a curve in a manner suggesting the existence of a finger, there is yet another test that needs to be done. The bar might be too small for the curve and vice versa. The widthDiff f(g) is given by the formula:

\[
\text{widthDiff}(f,g) = \sum_{r} \sum_{p} \text{intersects}(r, p)
\]

where \( \text{intersects}(r,p) \) gives the absolute difference between the width of the bar r and the diameter of the curve p.

If the score obtained from (2) is large enough for two groups f and g, they are merged into the smallest rectangle that can contain rectangles representing both f and g.

Step 150: Hand Tracking

In this step, the novel tracking technique of the present invention is applied to hand tracking. This technique makes use of motion information that can be obtained from a video. FIG. 2c is a flow diagram showing the steps in the track hand step 150, at each frame, a tracker is initialized in step 151 at all the hand locations found in the merged group step 140 for that specific frame. In step 152 that candidate hand is then tracked forward and backward in time. The location data from step 152 is record in step 153 to obtain a complete trajectory throughout the whole video in step 156 using a mean shift tracker described by Comaniciu and Meer which is initialized at the detection rectangle.

Assume this trajectory is called \( T(f,h) \), where f denotes the frame number the hand was found in, and h denotes the hand ID in that frame. Then this trajectory is composed of so called trajectory elements \( T(f,h,m) \):

\[
T(f,h,m)=T(f,h,1), \ldots , T(f,h,m), \ldots , T(f,h,M)
\] (5)

where M is the number of frames in the video and \( 1 \leq m \leq M \). Each trajectory element represents an \((x,y)\) location of the image.

It is assumed that the trajectories of the correct hands will be consistent throughout the video. False positives will not appear in all the frames; those that do appear will be static or in the background. Using this fact, in step 154 similar trajectories are merged, increasing their weight and choosing the trajectories that have the highest weights in step 156 to determine the hand trajectory in the end.

The average distance between two trajectories is defined as:

\[
\frac{\sum_{m=1}^{M} T_1(m)-T_2(m)}{M}
\]

where \( T_1(m) \) denotes the \((x,y)\) location of the trajectory at frame m, and the difference operator is the Euclidian distance between two points and M is defined to be the number of frames. Two trajectory elements \( T_1(j) \) and \( T_2(j) \) are called sufficiently close if:

\[
||T_1(j)-T_2(j)|| \leq \epsilon
\]

where \( \epsilon \) is the threshold for determining similarity.
where $i$ is a user defined threshold. In this example, $i$ is set to 5 pixels. Two trajectories are merged if 50% of their respective trajectory elements are sufficiently close. When two trajectories are merged in step 154, the resulting trajectory $T_n$ is:

$$T_n = \frac{T_1(1) - T_2(1)}{2} \ldots \frac{T_1(M) - T_2(M)}{2} \quad (8)$$

In step 154 when two trajectories are merged, their weights are added and this value becomes the weight of the new trajectory. Correct trajectories tend to accumulate and at the end have very high weights as shown by the merged groups shown in FIG. 7. After finding the correct trajectories, hands are segmented out using the trajectory histogram. For each hand in the detection phase, bar and curve histograms are created and merged with a detection is marked. Since the histogram is built for each hand in the detection phase, when merging trajectories, the respective histograms are merged to end up with a trajectory histogram in step 155. Using the center of the hand location as the seed, the hands are successfully extracted and separated at a scene in step 157 using a connected components algorithm.

If background subtraction is used, the largest connected component that is left and moving in the scene would be marked as the object. From that point, it is trivial to establish which hand is holding the object at a given time by comparing the velocity vectors of the objects to that of the detected hands in the scene. A velocity vector of an object between frames is found by subtracting the initial location from the final location to obtain the distance the hand traveled and dividing the distance by the number of frames elapsed to traverse this distance.

As an example, to understand the action of an object in a hand given to another (exchange), the steps shown in the Flow diagram of FIG. 2d are used. At step 162, the video is divided into intervals of 5 frames or longer. In step 164, for each interval, the object velocity is determined and in step 166 the object velocity is compared to each hand velocity to see which hand is more likely to carry the object. Then, in step 168, if at one point the object changes hands, in step 170 it is marked as an exchange. The present invention provides a tree of actions that can be determined in this framework. When two hands interact in, either the two hands are shaking or an object is being exchanged. When on hand is detected, the hand may or may not include an object. For the hand without an object, the choices are either move the object or pick up an object. Once the object is picked up, it becomes a hand with an object. For the hand with the object, the object can be moved or the object can be put down which moves the tracking back to being a hand without an object. The object is tracked from frame to frame following the flow diagram shown in FIG. 8.

The methods and systems according to the present invention were tested on ten (10) videos producing good results. Among the ten tested sequences, five were simple hand detection and tracking, while the rest were hands exchanging objects. In the five simple tracking cases, the hands were successfully detected and tracked through the sequences. No initialization or re-initialization was required during the detection process.

The results for the exchange scenes were also satisfactory. For the correct detection of an object exchange taking place at the scene, simple background subtraction is required to identify the object. In two out of five sequences, there were rare cases where the hand and object were mislabeled. The reason for this phenomenon in the sample cases is the similarity of the color distribution of the hand and the object.

The methods and systems of the preferred embodiment were also tested in dimly lit scenes, where the hand contours and details are not clearly visible. The loss of detail due to poor illumination affects the detection rate by reducing the probability of the hand primitives being found. Although this was the outcome in the experiments, there was little or no degradation in the tracking quality. Even if the tracker fails initiating from a single frame, other trackers that initiated from different frames compensated for the incorrect tracking.

In summary, the present invention provides novel tracking methods and systems with application to the hand tracking problem. Since the human hand has many degrees of freedom, it has been difficult to apply common detection and recognition techniques to hands. Instead, we detect what we term “finger primitives” and group these primitives to obtain the hands in the scene. Decision trees provide refinement of the grouping into hands. In the example shown and described, results with two primitive classes are illustrated.

The present tracking technique is not only a tracking application, but also can be used as an auxiliary system to eliminate false positives for weak detectors like the hand detector described in this system. The result of the tracking is used to detect the missed detections and to eliminate false positives.

This type of tracking is not limited to hand tracking. In fact any articulated object that can be identified by the detector is a candidate. Moreover, once the system gathers enough information about the hand, it is most likely to allow, it can extract the hand, therefore eliminating false positives. A proposed extension to the system is using a specialized tracking for the hand, instead of using mean shift. Although mean shift is superior to most of the generic tracking systems, its performance is low when applied to an articulated object. The information that is gathered in the detection steps can be utilized to improve the tracking.

While the invention has been described, disclosed, illustrated and shown in various terms of certain embodiments or modifications which it has presumed in practice, the scope of the invention is not intended to be, nor should it be deemed to be, limited thereby and such other modifications or embodiments as may be suggested by the teachings herein are particularly reserved especially as they fall within the breadth and scope of the claims here appended.

We claim:

1. A method for detection and tracking of a moving object comprises the steps of:
   - detecting at least two primitive classes for identifying the moving object within plural frames of video to generate plural primitive candidates for each class by detecting plural primitive candidates for each of the at least two primitive classes;
   - extracting a primitive model of each of the at least two primitive classes;
   - superimposing the primitive model on each primitive candidate; and
   - determining which primitive candidate is a primitive;
   - grouping each detected primitive candidates in each class separately according to a decision tree corresponding to each primitive class;
   - training the decision tree corresponding to the primitive class;
   - merging primitive groups for each class of primitives to identify the object being tracked; and
   - tracking the moving object through the plural frames of video according to the merged primitive groups.
2. The method of claim 1, wherein the grouping step comprises the steps of:

- creating one group for each detected primitive in each class to generate plural groups of primitive candidates for each of the at least two primitive classes using the decision tree that is trained corresponding to the primitive class;
- determining group membership of each of the plural groups from the plural primitive candidates detected in the detecting step;
- voting to eliminate primitive groups based on voting results; and
- eliminating subset groups and small groups to reduce false primitive positives, wherein the grouping step yields primitive candidate locations for each primitive class.

3. The method of claim 2, wherein the creating one group for each detected primitive candidate step comprises the step of:

- forming N groups when there are N primitives detected for a given primitive class, wherein each detected primitive is a representative of a different primitive group.

4. The method of claim 1, wherein the merging step comprises the steps of:

- merging primitive groups of a first class with primitive groups of a second class; and
- determining which merged first and second class groups include the object.

5. The method of claim 1, wherein the tracking the object through the plural frames of video comprises the steps of:

- tracking the moving object forward and backward through each of plural sets of video frames to obtain plural trajectories of the object;
- recording each of the plural trajectories;
- merging the plural trajectories;
- assigning weights to each of the merged trajectories;
- generating a trajectory histogram from the merged trajectories; and
- extracting and separating the moving object from the trajectory histogram.

6. A system for identifying and tracking a human hand in a video comprising:

- a video device for generating plural frames of video;
- a processor for executing instructions;
- a set of detecting instructions for detecting the human hand in the video comprising:
  - a parallel line set of instructions for detecting parallel line primitives;
  - a curved fingertip set of instructions for detecting curved fingertip primitives;
  - a grouping set of instructions for grouping the plural detected parallel line primitives and plural curved fingertip primitives according to a parallel line decision tree and a curved fingertip decision tree, respectively; and
  - a merging set of instructions for merging the parallel line groups and the curved fingertip groups to identify candidate hands; and
- a set of tracking instructions for tracking the human hand through the plural frames of video.

7. The system of claim 6, wherein the parallel line set of instructions for detecting parallel lines comprises:

- a first subset of parallel line instructions for detecting plural parallel lines in an image;
- a second subset of parallel line instructions for determining if two of the plural lines are parallel to generate plural pairs of parallel lines;
- a third subset of parallel line instructions for determining plural properties of each pair of parallel lines; and
- a fourth subset of parallel line instructions for eliminating pairs of parallel lines that are not candidates for fingers according to the plural properties.

8. The system of claim 6, wherein the curved fingertip set of instructions for detecting curved fingertips comprises:

- a first subset of curved fingertip instructions for extracting a model edge from a fingertip model;
- a second subset of curved fingertip instructions for grouping together neighboring pixels having a same orientation for each fingertip group;
- a third subset of curved fingertip instructions for marking a middle point of each curved fingertip group as a fingertip top candidate;
- a fourth subset of curved fingertip instructions for superimposing the fingertip model on each fingertip top candidate in the image;
- a fifth subset of curved fingertip instructions for scaling each fingertip model; and
- a sixth subset of curved fingertip instructions for matching each fingertip top candidate to the scaled fingertip model and a seventh subset of curved fingertip instructions for marking each fingertip top candidate as a fingertip when the matching to the superimposed fingertip model surpasses a threshold.

9. The system of claim 6, wherein the grouping set of instructions for grouping the plural detected parallel lines and plural curved fingertips comprises:

- a first subset of grouping instructions for creating one group for each detected parallel line and each detected curved fingertip to generate plural parallel line groups and plural curved fingertip groups;
- a second subset of grouping instruction for determining group membership of each of the plural parallel line groups and plural curved fingertip groups according to the parallel line decision tree and the curved fingertip decision tree, respectively;
- a third subset of grouping instruction for voting to eliminate parallel line groups and curved fingertip groups based on voting results; and
- a fourth subset of grouping instruction for eliminating subset groups and small groups to reduce false primitive positives, wherein the grouping step yields parallel line candidate locations and curved fingertip candidate locations.

10. The system of claim 6, wherein the merging set of instructions for merging the parallel line group and the curved fingertip groups to identify candidate hands comprises:

- a first subset of merging instructions for detecting intersections between the merged parallel line groups and curved fingertip groups; and
- a second subset of merging instructions for determining which intersecting parallel line groups and fingertip groups are finger candidates.

11. The system of claim 6, wherein the tracking set of instructions for tracking the detected human hand comprises:

- a first subset of tracking instruction for tracking the human hand forward and backward through each of plural sets of video frames to obtain plural trajectories of the human hand;
- a second subset of tracking instruction for recording each of the plural trajectories; and
- a third subset of tracking instruction for merging the plural trajectories.
a fourth subset of tracking instruction for assigning a
weight to each of the merged trajectories;
a fifth subset of tracking instruction for generating a tra­
jectory histogram from the weighted trajectories; and
a sixth subset of tracking instruction for extracting and
separating the human hands from the trajectory histo­
gram.
12. The system of claim 6, further comprising:
a set of object detection instructions for detecting an object
carried in one of the detected human hands.
13. The system of claim 12, further comprising:
a set of object tracking instructions for tracking a transfer
of the detected object from one detected human hand to
another detected human hand.
14. A method for detection and tracking of moving hands in
a video comprises the steps of:
detecting parallel line primitives in plural frames of the
video to generate plural parallel line primitives by
detecting plural lines in an image;
determining when two of the plural lines are parallel to
generate plural pairs of parallel lines;
determining plural properties of each pair of parallel
lines; and
eliminating pairs of parallel lines that are not candidates
for fingers according to the properties;
detecting curved fingertip primitives in plural frames of the
video to generate plural curved fingertip primitives;
grouping each parallel line primitive according to a par­
nallel line decision tree;
grouping each curved fingertip primitive according to a
curved fingertip decision tree; and
merging parallel line groups and curved fingertip groups to
identify the moving hand; and
tracking the moving hand through the plural frames of
video according to the merged primitive groups.
15. The method of claim 14, wherein the detecting curved
fingertip primitives step comprises the step of:
extracting a model edge from a fingertip model;
grouping together neighboring pixels having a same direc­
tional orientation for each curved fingertip group;
marking a middle point of each pixel group as a fingertip
top candidate;
superimposing the fingertip model on each fingertip top
candidate in the image;
scaling each fingertip model along at least one fingertip
model dimension;
comparing each fingertip top candidate to the scaled fing­
tertip model; and
marking each fingertip top candidate as a fingertip when
the comparison of the scaled superimposed fingertip
model with the fingertip top candidate surpasses a
threshold.
16. The method of claim 14, wherein the merging step
further comprises the steps of:
detecting intersections between the merged parallel line
groups and merged curved fingertip groups; and
determining which intersecting parallel line groups and
curved fingertip groups are a finger.
17. The method of claim 14, further comprising the steps of:
detecting an object in one of the hands;
dividing the video into intervals of 5 frames or longer;
determining the object velocity during each interval;
comparing the object velocity to each hand velocity to
determine which hand is more likely to carry the object;
and
marking one of the 5 or more frames as an exchange of the
object when the object changes hands in the frame.
18. The method of claim 17, wherein the tracking step
comprises the step of:
tracking the object forward and backward through each of
plural sets of video frames to obtain plural trajectories of
the object;
recording each of the plural trajectories;
merging the plural trajectories;
assigning weights to each of the merged trajectories;
generating a trajectory histogram from the merged trajec­
tories; and
extracting and separating the hands from the trajectory
histogram.
19. The method of claim 18, further comprising the steps of:
(a) determining how many hands are detected in each set of
video frames;
(b) when two hands are detected,
determining if the two hands interact;
(c) when two hands are interacting,
determining if the two hands are shaking; and
determining if an object is detected or object is being
exchanged;
(d) when one hand is detected,
determining if the one detected hand includes the object;
(d) when the object is detected with the hand;
determining if the hand is moving the object;
determining if the hand is putting the object down; and
repeating step (d) when the object is put down;
(e) when the object is not detected with the hand;
and
determining when the object is picked up by the hand; and
and
going to step (d) when the object is picked up;
repeating steps (a) through (e) for a next set of video
frames.
* * * * *