

METHODS AND SYSTEMS FOR USING CONFIDENCE METRICS FOR LEFT/RIGHT HAND IDENTIFICATION

TECHNICAL FIELD

[0001] The present invention generally relates to computer imaging, and more particularly relates to methods and systems for using confidence metrics for left/right hand identification.

BACKGROUND OF THE DISCLOSURE

[0002] Automation has become a key component in today's manufacturing processes. However, many tasks, especially on assembly lines, are still performed by humans due to their complexity and the need for precision. In recent times, the use of video streams for monitoring and analyzing assembly line processes is a common practice. These video streams can provide valuable data for improving efficiency and productivity.

[0003] In a video stream, accurately distinguishing between left hands and right hands is a challenging task due to the possibility of a hand going outside of the camera view and then re-entering, due to occlusions such as hands blocked by a head or other object or by overlap of the hands and due to fast movements of the hands.

[0004] As certain tasks are performed by one specific hand, misidentification of a hand could lead to incorrect movement patterns which could affect an analysis of these patterns and, thus, lead to incorrect measurement of cycle time and other analysis. In addition, the misidentification of a hand could yield incorrect hand trajectories, which would significantly impact and/or degrade trajectory-based analysis.

[0005] There exist several methods for left hand/right hand identification such as hand position-based tracking and machine learning models. However, the existing methods are either computationally intensive or each come with their own set of limitations. Such conventional methods include hand position-based tracking methods, using traditional machine learning models to detect left hands and right hands, image feature-based tracking, and hand landmark detection.

[0006] While hand position-based tracking methods can be used for left hand/right hand identification, if a hand ID is wrongly assigned, it will continue to be tracked incorrectly within a region of interest (ROI) and there is no mechanism to correct this misassignment while the hand remains within the ROI. Thus, hand position-based tracking methods are prone to errors and less robust than other hand position-based tracking methods.

[0007] Use of traditional machine learning models to detect left or right hand may lead to poor results as such models might struggle with various hand poses, hands covered by different colored gloves, uneven artificial lighting, and occlusion by objects and safety equipment. In addition, such machine learning models are computationally expensive.

[0008] Image feature-based tracking methods are computationally expensive and are prone to similarity of objects leading to wrong hand ID assignment. And hand landmark detection methods have difficulty with complex hand poses and the occlusion of a partial hand can lead to poor detection. In addition, hand landmark detection methods are also computationally expensive.

[0009] Thus, what is needed is methods and systems for determining left and right hands that is robust, reliable and computationally lightweight. Furthermore, other desirable features and characteristics will become apparent from the subsequent

detailed description and the appended claims, taken in conjunction with the accompanying drawings and this background of the disclosure.

SUMMARY

[0010] According to at least one embodiment of the present invention, a method for left/right hand determination in a frame within a plurality of frames of a video data stream is provided. The method includes detecting one or more hands in the frame, calculating an inter-frame distance of each of the one or more hands in the frame to a detected hand position in a previous frame of the plurality of frames, and assigning a hand ID to each of the one or more hands within the region of interest in the frame based on the calculated inter-frame distance. The method further includes computing a confidence metric based on a combination of positions of the detected hands in the frame and determining whether each of the one or more hands detected in the frame is a left hand or a right hand in response to the confidence metric and the assigned hand ID.

[0011] According to another embodiment of the present invention, a system for left/right hand determination in a frame within a plurality of frames of a video data stream is provided. The system includes an image capture device coupled to a computing device. The image capture device is configured to acquire video information. The computing device is configured to receive the video information as a video data stream and includes a hand detection unit, a confidence metric computation unit, and a left hand/right hand assignment unit. The hand detection unit is configured to detect one or more hands in a frame of a plurality of frames in the video data stream and calculate an inter-frame distance of each of the one or more hands in the frame to a detected hand position in a previous frame of the plurality of frames. The hand

detection unit is also configured to assign a hand ID to each of the one or more hands in the frame based on the calculated inter-frame distance. The confidence metric computation unit is configured to compute a confidence metric based on a combination of positions of the one or more hands detected in the frame. And the left hand/right hand assignment unit is configured to determine whether each of the one or more hands detected in the frame is a left hand or a right hand in response to the confidence metric and the assigned hand ID.

[0012] And according to yet another embodiment, a computer readable medium has instructions stored thereon for a processing means to perform left/right hand determination in a frame within a plurality of frames of a video data stream using a method which includes detecting one or more hands in the frame, calculating an inter-frame distance of each of the one or more hands in the frame to a detected hand position in a previous frame of the plurality of frames, assigning a hand ID to each of the one or more hands within the region of interest in the frame based on the calculated inter-frame distance, computing a confidence metric based on a combination of positions of the detected hands in the frame, and determining whether each of the one or more hands detected in the frame is a left hand or a right hand in response to the confidence metric and the assigned hand ID.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] The accompanying figures, where like reference numerals refer to identical or functionally similar elements throughout the separate views and which together with the detailed description below are incorporated in and form part of the specification,

serve to illustrate various embodiments and to explain various principles and advantages in accordance with a present embodiment.

[0014] FIG. 1 depicts typical hand identification (ID) methodology.

[0015] FIG. 2, comprising FIGs. 2A and 2B, depicts examples of instances of more and less confidence of hand ID assignment in accordance with the present embodiments, wherein FIG. 2A depicts an example of high confidence of correct hand ID assignment and FIG. 2B depicts examples of low confidence of correct hand ID assignment.

[0016] FIG. 3 depicts four frames of a data stream in accordance with the present embodiments in a scenario where both hands are moved out of the camera view and then one hand and later the other hand come back into the camera view.

[0017] FIG. 4 depicts a block diagram of realtime hand tracking system in accordance with the present embodiments.

[0018] FIG. 5 depicts a flowchart of operation of the realtime hand tracking system of FIG. 4 in accordance with the present embodiments.

[0019] FIG. 6 depicts a block diagram of a realtime hand tracking system with a pre-calculation unit in accordance with the present embodiments.

[0020] FIG. 7 depicts flowchart of operation of the realtime hand tracking system of FIG. 6 with the pre-calculation unit in accordance with the present embodiments.

[0021] FIG. 8 depicts a block diagram of a hand tracking system with batch processing in accordance with the present embodiments.

[0022] FIG. 9 depicts a flowchart of operation of the hand ID assignment unit of the hand tracking system with batch processing of FIG. 8 in accordance with present embodiments.

[0023] FIG. 10 depicts a block diagram of a combination of batch and realtime hand tracking system in accordance with the present embodiments.

[0024] FIG. 11, comprising FIGs. 11A and 11B, depicts a first method for computing the confidence metric based on the difference between the x-coordinate of detected hands, the y-coordinate of detected hands, and the distance between the hands and a body position in accordance with the present embodiments, wherein FIG. 11A depicts a high confidence case and FIG. 11B depicts a low confidence case.

[0025] FIG. 12, comprising FIGs. 12A, 12B and 12C, depicts confidence maps of parameters used to compute the confidence metric by the first method in accordance with the present embodiments.

[0026] FIG. 13 depicts a flowchart of computing the confidence metric by the first method in accordance with the present embodiments.

[0027] FIG. 14, comprising FIGs. 14A to 14D, depicts illustrations of initial three steps of a second method which uses a range of motion of a hand to compute the confidence metric in accordance with the present embodiments, wherein FIG. 14A depicts an initial joint detection step in the second method, FIG. 14B depicts a vector formation step in the second method, and FIGs. 14C and 14D depict range of motion estimation steps in the second method.

[0028] FIG. 15, comprising FIGs. 15A to 15D, depicts illustrations of a hand confidence map determination step in the second method to compute the confidence metric in accordance with the present embodiments, wherein FIG. 15A depicts a confidence map determined from a range of motion of a hand, FIG. 15B depicts a left hand range of motion forming a confidence map on a workbench, FIG. 15C depicts a right hand range of motion forming a confidence map on the workbench and FIG. 15D

depicts a confident inter hand map and a non-confident inter hand map on the workbench.

[0029] FIG. 16, comprising FIGs. 16A and 16B, depicts illustrations of further steps of the second method when computing the confidence metric in accordance with the present embodiments, wherein FIG. 16A depicts a person movement simulation step and FIG. 16B depicts a multi-person workbench step.

[0030] FIG. 17, comprising FIGs. 17A, 17B and 17C, depicts a confidence metric computation step of the second method in accordance with present embodiments, wherein FIG. 17A depicts a first step in the confidence metric computation, FIG. 17B depicts a second step in the confidence metric computation, and FIG. 17C depicts a third step in the confidence metric computation.

[0031] FIG. 18, comprising FIGs. 18A and 18B, depicts initial steps in a third method for confidence metric calculation in accordance with the present embodiments, wherein FIG. 18A depicts a joint detection step and FIG. 18B depicts a vector length calculation step.

[0032] FIG. 19 depicts two possible ways to draw bounding boxes for a same hand position in accordance with the present embodiments, wherein a rotated bounding box provides more information for estimation in further steps.

[0033] FIG. 20 depicts estimation of possible wrist position, elbow location and shoulder location by the third method in accordance with the present embodiments.

[0034] FIG. 21, comprising FIGs. 21A, 21B and 21C, depicts two scenarios for confidence metric computation by the third method in accordance with the present embodiments, wherein FIG. 21A depicts a first scenario and FIGs. 21B and 21C depict two steps of a second scenario.

[0035] FIG. 22, comprising FIGs. 22A and 22B, depicts use of a grid for computing a lookup table of confidence metrics for the fourth method of confidence metric computation in accordance with the present embodiments, wherein FIG. 22A depicts a first step and FIG. 22B depicts a second step.

[0036] FIG. 23 depicts identifying detected hands plotted to cells of the grid utilized in FIGs. 22A and 22B in accordance with the present embodiments.

[0037] FIG. 24, comprising FIGs. 24A and 24B, depicts a visualization of the confidence metric computation of the fourth method in accordance with the present embodiments, wherein FIG. 24A depicts a probability of a correct position occurrence and FIG. 24B depicts a probability of an opposite position occurrence.

[0038] FIG. 25, comprising FIGs. 25A, 25B and 25C, depicts an example of how tracklets are generated in accordance with the present embodiments, wherein FIG. 25A depicts the elements of the hand ID assignment unit of the hand tracking system of FIG. 8 with batch processing and FIGs. 25B and 25C depict a scenario of tracklet generation.

[0039] And FIG. 26, comprising FIGs. 26A and 26B, depicts zone creation and interpolation in accordance with the present embodiments, wherein FIG. 26A depicts zone creation and tracklet connection and FIG. 26B depicts hand detection interpolation.

[0040] Skilled artisans will appreciate that elements in the figures are illustrated for simplicity and clarity and have not necessarily been depicted to scale. In addition, same reference numerals in different figures represent the same feature or element.

DETAILED DESCRIPTION

[0041] The following detailed description is merely exemplary in nature and is not intended to limit the invention or the application and uses of the invention. Furthermore, there is no intention to be bound by any theory presented in the preceding background

of the invention or the following detailed description. It is the intent of the present embodiments to present methods and systems which use a combination of detected hand positions to judge a confidence of a left-hand identification or a right-hand identification without using a computationally-intensive algorithm. The proposed scheme uses left hand and right hand positions in a camera view to accurately distinguish between them. In accordance with present embodiments, if detected hands are sufficiently distanced from each other in a camera view, such as at a distance larger than what is observed when the hands are crossed, it is inferred that the leftmost detection corresponds to the left hand and the rightmost detection corresponds to the right hand. Thus, the methods and systems in accordance with the present embodiments are computationally lightweight. In addition, the scheme in accordance with the present embodiments can advantageously be used as an addition to hand position-based tracking methods, thereby improving accuracy and efficiency and beneficially enabling correction of hand ID at any stage of the hand position-based tracking method.

[0042] Referring to FIG. 1, imaging information a camera view 100 is provided to a hand position tracking method 110 which is used to identify, in the camera view, the right hand and the left hand. Tracking methods that are deployed on edge devices are often constrained by limited computational capabilities, so it is preferred that the tracking method 110 be computationally lightweight when assigning a unique hand ID for each hand in the camera view, yet correctly identify the correct hand (i.e., provide the correct hand ID assignment). Incorrect hand assignment, such as assigning a left hand a right hand ID 120 and assigning a right hand a left hand ID 125 is a problem with many hand position tracking methods.

[0043] Referring to FIG. 2A, illustration 200 depicts an example of high confidence of correct hand ID assignment in accordance with the present embodiments. As a hand

210 is detected a large distance 215 from a hand 220, a correct hand identification of the hand 210 on the left as the left hand and the hand 220 on the right as the right hand can be made with high confidence as the large distance 215 indicates that it would be extremely difficult for the hands to be crossed.

[0044] Referring to FIG. 2B, illustrations 230 and 260 depict examples of low confidence of correct hand ID assignment in accordance with the present embodiments. In illustration 230, a hand 240 is detected a small distance 245 from a hand 250. Similarly, in illustration 260, a hand 270 is detected a small distance 275 from a hand 280. When one hand is detected at a small distance from the other hand, there is a chance the hands might be crossed. For example, in the illustration 230, the hands 240, 250 are crossed and in the illustration 260, the hands 270, 280 are not crossed. Therefore, the small distance 245, 275 between the hands indicates a low confidence of correct hand ID assignment.

[0045] The illustration 300 in FIG. 3 depicts four frames 310, 320, 330, 340 of a data stream where a camera is filming an operator's hands working on a workbench in the camera's view in accordance with the present embodiments. In the first frame 310, both hands 302, 304 are in the camera view. In the second frame 320, both hands 302, 304 are moved out of the camera view as the operator fetches a thing located out of the camera's view. Then one hand is moved back into the camera view in the third frame 330. It is unclear which hand is in the camera view in the frame 330 and the hand ID cannot be determined with only the positional information. If such position-based determination would be made, it would be incorrect because, even though the left hand comes back in view, it is closest to the position where the right hand was before the two hands were moved out of the camera's view. In the fourth frame 340, the other hand comes back into the camera view and it can be determined that the previously appearing

hand in the third frame is hand 302. Thus, it can be seen that the leftmost hand in the fourth frame is hand 302 and the rightmost hand is hand 304.

[0046] Four variations of hand tracking systems and methods with confidence metric computation units in accordance with the present embodiments will be presented. These include a realtime hand tracking system and method, a realtime hand tracking system and method with a pre-calculation unit, a hand tracking system with non-realtime processing, and a combination realtime and non-realtime hand tracking system and method.

[0047] First, referring to FIG. 4, a block diagram 400 depicts the realtime hand tracking system 410 in accordance with a first variation of the present embodiments. The system 410 operates in realtime, providing immediate feedback based on hand movements. The system 410 includes a camera 420 and a computing device 430. Those skilled in the art will recognize that the computing device could be a server, a single processor, or multiple computers communicatively coupled together. The camera 420 records a video data stream which is fed to the computing device 430. An image processing unit 432 of the computing device 430 receives the video data stream from the camera 420 and processes the data stream frame-by-frame into a format suitable for further processing. Each frame is forwarded to a hand detection unit 434 which is responsible for identifying and locating hands in the processed images by using object detection for bounding box detection to isolate the region of interest (i.e., hands). This information is provided to a confidence metric computation unit 436 of the computing device 430 which calculates, in accordance with the present embodiments, a confidence metric for hand ID assignment of detected hands. The confidence metric represents whether the hands can be distinguished left or right with high confidence or low confidence. In response to the confidence metric, a hand ID determination unit 438

determines a unique identifier (ID) for each detected hand. The ID is used to track the same hand across different frames in the video feed. A hand ID assignment unit 440 assigns the determined ID to each detected hand. The assignment is based on various factors, ensuring consistent tracking of each hand. A visualization unit 442 of the computing device 430 is responsible for visualizing the output in accordance with the hand ID assignments, such as overlaying bounding boxes, hand trajectories on the original video feed, or displaying the processed images with detected hands and their IDs.

[0048] FIG. 5 depicts a flowchart 500 of operation of the realtime hand tracking system 410 in accordance with the first variation of the present embodiments. Processing starts by the image processing unit 432 determining if a frame is read 502. If a frame is read 502, the hand detection unit 434 examines the frame to detect a hand 504.

[0049] If the number of detected hands is not equal to a defined number of hands (i.e., twice the number of persons in the frame) 506 and no previous position is available 508, there is not enough information for the confidence metric computation unit to determine a confidence metric, so processing returns to the image processing unit 432 reading a subsequent frame 502.

[0050] When the number of hands is equal to the defined number of hands 506 and no previous position is available 510, then the frame being examined is a first frame in which one or both hands are detected, and the hand ID assignment unit 440 sorts the detected hands from left to right based on the x coordinates of the detected hands and assigns hand IDs 512 to the detected hands and processing returns to read the next frame 502.

[0051] When the number of detected hands is equal to the number of hands defined 506 and a previous position is available 510, then the hand detection unit 434 calculates the distance of each hand detected in the current frame to their previous position (i.e., an inter-frame distance) and assigns a hand ID of a closest hand detection 514. Next, the confidence metric computation unit 436 computes the confidence metric 516 and forward the confidence metric to the hand ID determination unit 438 to determine if the confidence metric is greater than a threshold confidence metric 518 or not.

[0052] When the number of detected hands is not equal to the number of hands defined 506 yet a previous position is available 508, the missing detections are imputed from their previous positions 520 and the hand detection unit 434 calculates the distance of each hand detected in the current frame to their previous position and assigns a hand ID of a closest hand detection 514. The confidence metric computation unit 436 then computes the confidence metric 516 and forward the confidence metric to the hand ID determination unit 438 to determine if the confidence metric is greater than a threshold confidence metric 518 or not.

[0053] If the hand ID determination unit 438 determines that the confidence metric is greater than a threshold confidence metric 518, the hand ID assignment unit 440 sorts the detected hands from left to right based on the x coordinates of the detected hands and updates the hand IDs 512 of the detected hands and processing returns to read the next frame 502. If the hand ID determination unit 438 determines that the confidence metric is not greater than a threshold confidence metric 518, processing returns to read the next frame without updating the hand IDs.

[0054] Next, referring to FIG. 6, a block diagram 600 depicts the realtime hand tracking system 610 with a pre-calculation unit 624 in accordance with a second

variation of the present embodiments. This system also functions in realtime but includes the pre-calculation unit 624 that processes data prior to the actual tracking.

[0055] The system includes the camera 420 and a computing device 620. An image processing unit 622 of the computing device 620 receives the data stream from the camera 420 and forwards the data stream from the camera 420 to the pre-calculation unit 624 which processes the data stream to estimates confidence maps or confidence functions and stores the confidence maps or confidence functions in a confidence map/function storing unit 626 of a memory unit of the computing device 620. The memory unit in the computing device is a crucial component that stores data and instructions and provides the necessary storage capacity for the system to function, holding information temporarily for processing or permanently for retrieval later. The confidence maps stored and updated in the confidence map/function storing unit 626 of the memory unit of the computing device 620 may be later accessed for confidence metric computation.

[0056] The image processing unit 622 also receives the video data stream from the camera 420 and processes the data stream frame-by frame into a format suitable for further processing. Each frame is forwarded to a hand detection unit 434 which is responsible for identifying and locating hands in the processed images by using object detection for bounding box detection to isolate the region of interest (i.e., hands) to determine whether a hand is detected or no hands are detected. This information is provided to a confidence metric computation unit 628 of the computing device 620 which calculates, in accordance with the present embodiments, a confidence metric for hand ID assignment of detected hands. The confidence metric represents whether the hands can be distinguished left or right with high confidence or low confidence. The operation of the confidence metric unit 628 is similar to the confidence metric unit 436

of the computing device 430, but the confidence metric is computed based on the confidence map or function stored in the confidence map/function storing unit 626. The confidence metric unit 628 may also provide the confidence metric for updating the confidence map and the confidence function stored in the confidence map/function storing unit 626.

[0057] In response to the confidence metric, a hand ID determination unit 438 determines a unique identifier (ID) for each detected hand. The ID is used to track the same hand across different frames in the video feed. A hand ID assignment unit 440 assigns the determined ID to each detected hand. The assignment is based on various factors, ensuring consistent tracking of each hand. A visualization unit 442 of the computing device 620 is responsible for visualizing the output in accordance with the hand ID assignments, such as overlaying bounding boxes, hand trajectories on the original video feed, or displaying the processed images with detected hands and their IDs.

[0058] FIG. 7 depicts a flowchart 700 of operation of the realtime hand tracking system 610 with the pre-calculation unit 624 in accordance with the second variation of the present embodiments. The operation of the realtime hand tracking system 610 in the flowchart 700 is essentially the same as the operation of the realtime hand tracking system 410 in the flowchart 500 with two differences between the flowchart 500 and the flowchart 700. First, before the frame is read 502 by the image processing unit 432, confidence maps or functions are pre-calculated 710. Next, the confidence maps or functions are used in the confidence metric computation 720. By pre-calculating 710 the confidence maps or functions prior to the actual tracking and then using the confidence maps or functions for the confidence metric computation 720

enables quicker, more robust realtime hand ID assignment by the hand ID assignment unit 440.

[0059] Next, referring to FIG. 8, a block diagram 800 depicts the hand tracking system 810 with non-realtime processing in accordance with a third variation of the present embodiments. The system 810 utilizes non-realtime processing in accordance with this third variation of the present embodiments, thereby allowing for the handling of hand tracking data in groups rather than individually for a more robust performance. The system 810 includes the camera 420 and a computing device 820. The camera 420 records a video data stream which is fed to the computing device 820. The image processing unit 432 of the computing device 820 receives the video data stream from the camera 420 and processes the data stream frame-by-frame into a format suitable for further processing. Each frame is forwarded to the hand detection unit 434 which is responsible for identifying and locating hands in the processed images by using object detection for bounding box detection to isolate the region of interest (i.e., hands) to determine whether a hand is detected or no hands are detected. This information is provided to the confidence metric computation unit 436 of the computing device 820 which calculates, in accordance with the present embodiments, the confidence metric for hand ID assignment of detected hands. In response to the confidence metric, a hand ID determination unit 438 determines the unique identifier (ID) for each detected hand. A hand ID assignment unit 850 assigns the determined ID to each detected hand. The visualization unit 442 of the computing device 820 is responsible for visualizing the output in accordance with the hand ID assignments, such as overlaying bounding boxes, hand trajectories on the original video feed, or displaying the processed images with detected hands and their IDs.

[0060] In accordance with this third variation of the present embodiments, the hand ID assignment unit 850 is modified to accommodate non-realtime processing and includes a tracklet generator unit 852, a zone creation unit 854, a tracklet connector unit 856 and an interpolation unit 858. The tracklet generator unit 852 is responsible for creating tracklets, which are short sequences of object locations over time. Each tracklet entirely consists of either hands with high confident metrics or low confident metrics. The tracklet generator unit 852 plays a crucial role in object tracking. The zone creation unit 854 is responsible for creating zones of tracklets. A zone is created every time at least one of the hand tracklets is a non-confident tracklet and will be ended when all the hand tracklets are confident. The tracklet connector unit 856 is responsible for linking tracklets based on time and spatial information to form complete tracks within a zone. The tracklet connector unit 856 ensures continuity in object tracking, even when the object is temporarily occluded or out of the frame. For any missing detections in a time sequence, the interpolation unit 858 fills in gaps by interpolating between a previous detection and a next detection. The interpolation is done by using the previous detection or calculating the midpoint between two detections, thereby maintaining the accuracy of tracking. The operation of hand ID assignment unit 850 is shown in a flowchart 900 of FIG. 9.

[0061] The flowchart 900 depicts the operation of the hand ID assignment unit 850 in accordance with the third variation of the present embodiments. Processing first determines whether a hand detection is the last detection 902. If the hand detection is not the last detection 902, the tracklet generator unit 852 checks for eligible active tracklets 904. If there are no eligible active tracklets 904, the tracklet generator unit 852 creates a new tracklet 906. Then the active tracklets identified at step 904 or created at step 906 are grown 908 as follows: based on spatial and time information confident

tracklets are grown by joining confident detected hands and non-confident tracklets are grown by joining non-confident hands.

[0062] Once all the hands are detected 902 (i.e., after the last detection 902), non-realtime processing is performed by the hand ID assignment unit 850. The tracklets are processed one-by-one until the last tracklet is detected 910. The tracklet processing includes the zone creation unit 854 creating zones 912. A zone of tracklets is started when at least one hand becomes non-confident and continues until all hands are confident again. When a zone is created 914, the tracklet connector unit 856 joins the tracklets within the zone using a shortest path approach, updating the non-confident tracklet IDs to connected confident tracklet IDs. After all of the tracklets are processed 910, the interpolation unit 858 interpolates any missing hand detections using a previous or average position method. If a zone is not created 914, the processing returns to check the next tracklet is the last one 910.

[0063] Finally, referring to FIG. 10, a block diagram 1000 depicts the combination non-realtime and realtime hand tracking system 1010 in accordance with a fourth variation of the present embodiments. The system 1010 amalgamates the features of both non-realtime processing and realtime tracking, aiming to leverage the benefits of both systems. The system 1010 includes the camera 420 and a computing device 1020. The computing device includes the components of the realtime tracking system 410 where the output of the hand determination unit 438 is coupled to both the input of the hand ID assignment unit 440 of the realtime tracking system 410 and the input of the hand ID assignment unit 850 of the non-realtime processing system 810. While the output of the hand ID assignment unit 440 is provided to the visualization unit 442 for visualization in real time, the more accurate and robust output of the hand ID

assignment unit 850 is provided to a memory unit 1030 of the computing device 1020 to advantageously save for further processing.

[0064] For each of the four hand tracking system and method variations of the present embodiments, four alternative methods for the confidence metric computation unit 436, 628 to compute the confidence metric 516, 720 are presented as a key unit of and integrated into the various systems and methods in accordance with the present embodiments. A first method for computing the confidence metric 516, 720 is based on the difference between the x-coordinate of detected hands, the y-coordinate of detected hands, and the distance between the hands and a body position. A second method for computing the confidence metric 516, 720 fixes a body in one position, based on the position of the wrist, elbow and shoulder joints and the confidence maps are generated by a range of motion of the hands. This process is repeated for all possible body positions. The confidence maps are stored in the memory unit of the computing device 430, 620, 820, 1020 and accessed based on a detected/predefined body position in order to compute the confidence metric. A third method involves calculating lengths of forearms, humerus and clavicles and storing them in the memory unit. An inverse kinematics solver is then used to estimate possible elbow and shoulder positions based on constraints like wrist position, range of rotation, movement, and lengths of forearm, humerus and/or clavicle. Based on this information, the confidence metric is computed to check for a possibility of the hands being crossed. And a fourth method creates, in the pre-calculation unit 624, a grid within a region of interest and a confidence metric is computed for each possible combination of hand positions and stored in the memory unit of the computing device 430, 620, 820, 1020. Any of the four methods can be utilized to create a look up table with positions and confidence metrics. Based on the

position of a hand, the confidence metric can be retrieved from the table. The four methods will be explained in more detail hereinafter.

[0065] Referring to FIGs. 11A and 11B and illustrations 1100, 1150, the basis for the first method for computing the confidence metric 516, 720 based on the difference between the x-coordinate of detected hands, the y-coordinate of detected hands, and the distance between the hands and a body position in accordance with the present embodiments is depicted. Here the y-axis is defined downward, with the y-coordinate increasing as a hand moves downward. The first method is a formulation for computing the confidence metric by finding conditions that can clearly distinguish left and right hands based on a combination of detected hand positions. The illustration 1100 of FIG. 11A depicts a high confidence metric case where the probability of swapping left-hand and right-hand positions is low. In the illustration 1100, the difference of the x coordinates d_x 1110 is large meaning that both hand positions are separated. The difference of the y coordinates of both hands, d_y 1120, is small and the y coordinates of both hands are close to person body as indicated by arrow 1130.

[0066] The illustration 1150 of FIG. 11B depicts a low confidence metric case where the probability of swapping left-hand and right-hand positions is high. In the illustration 1150, the difference d_x 1160 of the x coordinates is small and the difference d_y 1170 of the y coordinates of both hands is not small.

[0067] The differences d_x and d_y are, thus, used to compute the confidence metric (i.e., the probability of hand ID assignment correctness). In addition, another condition that the number of detected hands is correct is also used as well as the closeness of the hands to the person as seen with the arrow 1130 in the illustration 1100. The position of the person may be predefined or determined separately. For example, the person

position can be determined by applying person detection, or derived from temporal statistics of hand position by combining a robust statistical method.

[0068] The confidence metric can be calculated by using functions reflecting the characteristics mentioned above as shown in the three graphs of FIGs. 12A, 12B and 12C with Equation 1.

$$Conf = p_1(d_x)p_2(d_y)p_3(y_{ave}) \quad (1)$$

[0069] Equation 1 is one example of a computation of the confidence metric by the first method in accordance with the present embodiments. Alternative scoring mechanisms for computing confidence metrics or confidence maps (which are generated in the pre-calculation unit 624 (FIG. 6)) can also be implemented such as shown in Equation 2 or Equation 3.

$$Conf = p_1(d_x)+p_2(d_y)+p_3(y_{ave}) \quad (2)$$

$$Conf = \ln(p_1(d_x)) + \ln(p_2(d_y)) + \ln(p_3(y_{ave})) \quad (3)$$

where the function $\ln(x)$ in Equation 3 denotes the natural logarithm.

[0070] Referring to FIG. 13, a flowchart 1300 depicts computation of the confidence metric 516, 720 by the first method in accordance with the present embodiments. At a first step 1302, the term $p_1(d_x)$ is computed based on the distance between x coordinates of the hand positions. Then, at a second step 1304, the term $p_2(d_y)$ is computed based on the distance between y coordinates of the hand positions. Next, at a third step 1306, the term $p_3(y_{ave})$ is computed based on the distance between an average of the coordinates of the hand positions. Finally, at a fourth step 1308, the confidence metric is calculated based on $p_1(d_x)$, $p_2(d_y)$ and $p_3(y_{ave})$ using Equation 1, Equation 2, Equation 3 or similar robust statistics or related mechanisms or schemes.

[0071] The second method for computing the confidence metric 516, 720 in accordance with the present embodiments fixes a body in one position and, based on

the position of the wrist, elbow and shoulder joints, the confidence maps are generated by a range of motion of the hands. This process is repeated for all possible body positions and the confidence maps are stored in the memory unit of the computing device 430, 620, 820, 1020 and accessed based on a detected body position in order to compute the confidence metric. In accordance with this second method, based on the range of motion of hand, hand confidence maps are estimated in the pre-calculation unit 624, stored in the confidence map/function storing unit 626 of the memory unit of the computing device 620, and used for confident metric computation by the confidence metric computation unit 628.

[0072] FIG. 14A depicts, in illustration 1400, an initial joint detection step in the second method. The initial joint detection step involves the detection of the elbow and wrist for the forearm joint and the shoulder and the elbow for the humerus joint. Image processing schemes such as image segmentation or key point detection techniques may be used for joint detection step.

[0073] FIG. 14B depicts, in illustrations 1420 a vector formation step in the second method, following the detection of the joints, which uses vectors to represent the forearm and humerus. The forearm is represented as a vector that starts at the elbow and ends at the wrist, while the humerus is represented as a vector that begins at the shoulder and concludes at the elbow.

[0074] FIGs. 14C and 14D depict, in illustrations 1440, 1470, a range of motion estimation step. In the illustration 1440, the range of motion of the shoulder joint is typically from 45 degrees of horizontal flexion to 135 degrees of horizontal extension. Based on the range of motion of the shoulder, a maximum and minimum probable elbow positions are estimated for each hand

[0075] Typically, the range of motion of the elbow joint is from 40 degrees to 180 degrees (full extension) of flexion as seen in the illustration 1470. Based on the range of motion of the elbow, a maximum and minimum probable wrist positions are estimated for each hand.

[0076] Illustrations 1500, 1520, 1540, 1560 in FIGs. 15A to 15D, respectively, depict visually a hand confidence map determination step in the second method to compute the confidence metric in accordance with the present embodiments. The illustration 1500 illustrates the maximum and minimum positions a hand can reach when flexing the shoulder joint and the elbow joint of a person within a range of interest in front of the person. Based on the maximum and minimum positions that a hand can reach within the range of interest, a confidence map 1510 for each hand is created. The illustration 1520 depicts a left hand range of motion with a confidence map on a workbench 1522, and the illustration 1540 depicts a right hand range of motion with a confidence map on the workbench 1522. From the confidence hand maps, a high confident map region 1570 and a non-confident map region 1565 are determined. Confident inter hand maps 1570 are non-intersected areas of hand confidence maps of a person and low confident inter hand maps 1565 are intersected areas of hand confidence maps of a person. Referring to the illustration 1560, a low confident inter hand map is delineated by the dashed line 1565 between the hands of the person while the confident inter hand map is all other regions 1570 within the range of motion of the left hand and the right hand on the workbench 1522.

[0077] FIGs. 16A and 16B depicts illustrations 1600, 1650, respectively, of further steps in the second method to compute the confidence metric in accordance with the present embodiments. The illustration 1600 depicts a person movement simulation, where the person position is simulated moving from a left-hand side 1610 of the grid

cell until the right-hand side 1620 of the grid cell to estimate the confidence map for every position. All the confidence maps are computed and stored in the memory unit of the computing device.

[0078] The illustration 1650 depicts an extension of the second method to a multi-person workbench 1660. Note that a non-confident intra hand confidence map 1660 is formed at intersected areas of hand confidence maps between different persons.

[0079] As a final step in the second method, the confidence metric is computed. Referring to an illustration 1700 of FIG. 17A, as a first step in the confidence metric computation, a region of interest, such as a workbench, is divided into a plurality of cells 1710 and a centroid 1720 of a bounding box 1722 of a detected person 1724 is determined. Then, a second step in the confidence metric computation as shown in an illustration 1730 of FIG. 17B includes determining, based on the centroid 1720 which respective confidence map to retrieve. As shown in the illustration 1730 of FIG. 17B, the centroid 1720 lines up with cell four (4) 1735. So, as depicted in an illustration 1760 of FIG. 17C, the confidence map is retrieved for cell four (4) 1735 and the confidence metric is computed. The confidence maps are processed using the first method in the pre-calculation unit 624 and stored in the confidence map/function storing unit 626 of the memory unit of the computing device.

[0080] In each frame, from the detection hand position, the hand confidence metric computation based on various parameters including a parameter p1 indicating where the hand position is located in the confidence map, a parameter p2 indicating in which confidence map the hand position is located, and a parameter p3 indicating how close the hand is to another confidence map. The confidence metric is then calculated by a score obtained through each parameter such as a calculation in accordance with Equation 4.

$$Conf = s_1(p_1) s_2(p_2) s_3(p_3) \quad (4)$$

The Equation 4 is one example of a computation of the confidence metric by this second method in accordance with the present embodiments. Alternative scoring mechanisms can also be implemented such as shown in Equations 5 and 6.

$$Conf = s_1(p_1) + s_2(p_2) + s_3(p_3) \quad (5)$$

$$Conf = \ln (s_1(p_1)) + \ln (s_2(p_2)) + \ln (s_3(p_3)) \quad (6)$$

[0081] The calculated confidence metrics for each cell are stored in a lookup table in the pre-calculation unit 624 and stored in the confidence map/function storing unit 626 of the memory unit of the computing device. An exemplary lookup table for cell-based confidence maps is shown in TABLE 1.

Cell	Map
1	Map 1
2	Map 2
3	Map 3
.	.
.	.
.	.
10	Map 10

TABLE 1

[0082] A third method involves calculating lengths of forearms, humerus and clavicles and storing the lengths in the memory unit. An inverse kinematics solver is then used to estimate possible elbow and shoulder positions based on constraints like wrist position, range of rotation, movement, and lengths of forearm, humerus and/or clavicle. Based on this information, the confidence metric is computed to check for a possibility of the hands being crossed.

[0083] The first steps of the third method are a joint detection step and a vector length calculation step which are both processed in the pre-calculation unit 624 to identify the length of each vector. An illustration 1800 in FIG. 18A depicts the initial joint detection step which involves the detection of the joints, which are detection of the elbow joint and the wrist joint for calculating a length of the forearm and the shoulder joint and the elbow joint for calculation of a length of the humerus.

[0084] Referring next to an illustration 1850 in FIG. 18B, the vector length calculation step is depicted. Following detection of the wrist, elbow and shoulder joints, the joints are used to represent the forearm, humerus, clavicle as vectors. The forearm is represented as a vector that starts at the elbow and ends at the wrist. The humerus is represented as a vector that begins at the shoulder and concludes at the elbow. And the clavicle is a connection between shoulders which is represented as a vector that begins at one shoulder and ends at the other shoulder. The length of vector is determined for each of the forearm vector, the humerus vector and the clavicle vector. In accordance with the present embodiments the length of both left and right forearm vectors and both left and right humerus vectors is assumed to be the same. The length of each of the vectors is a one time calculation performed in the pre-calculation unit 624.

[0085] Referring to FIG. 19, an illustration depicts a traditional bounded box 1910 for hand position detection and a rotated bounded box 1920 for hand detection in accordance with the third method. The object detection model used for hand detection in accordance with the third method is more robust as it contemplates that the bounded box 1920 for a hand position may be rotated by an angle θ from the traditional bounded box 1910 which limits possible movement of the hand.

[0086] Referring to FIG. 20, estimation of possible wrist position, elbow location and shoulder location is depicted in the illustration 2000. For every detected hand, a lower

edge midpoint (LEM) is considered as an estimated wrist joint position. An inverse kinematics solver is used to simulate all the possible combinations of elbow and shoulder joints, based on the estimated wrist position, a range of rotation of the wrist, the elbow and the shoulder joints, and fixed lengths of the forearm, the humerus and the clavicle vectors. This process is repeated, frame-by-frame.

[0087] For confidence metric computation by the third method in accordance with the present embodiments, if any one shoulder joint could reach both wrist joints from the simulated possible positions, then a possibility of hands being crossed exists. Only if both shoulder joints can reach one exclusive wrist joint, then the hands are considered not crossed.

[0088] Referring to FIG. 21A, an illustration 2100 depicts hands being detected at the positions as shown for a first scenario. Accordingly for this first scenario, the hands are considered not crossed because in all the possible combinations for shoulder joint j_3 , it could only reach one wrist joint j_1 and shoulder joint j_4 could only reach one wrist joint j_6 . Hence, the confidence metric will be 1 and the hands are considered not crossed.

[0089] Referring to FIGs. 21B and 21C, ~~a first and second step of a~~ in second scenario two possible ways to reach wrist joint from shoulder joint are depicted in illustrations 2130 and 2160, respectively. Hands are detected at the positions as shown in the illustrations 2130, 2160. In this second scenario, there is a chance the hands are crossed. In all the possible combinations for shoulder joint j_3 , it could reach wrist joint j_1 as shown in the illustration 2130 and it could reach another wrist joint j_6 as shown in the illustration 2160. Hence the confidence metric is 0 as there is a chance for the hands to be crossed. Even if j_4 could only reach one wrist joint j_6 , since the joint j_3 can reach both wrist joints j_1 and j_6 , the confidence metric is 0.

[0090] The fourth method creates, in the pre-calculation unit 624, a grid within a region of interest and a confidence metric is computed for each possible combination of hand positions and stored in the memory unit of the computing device 430, 620, 820, 1020. Any method can be utilized to create a look up table with positions and confidence metrics. Based on the position of a hand, the confidence metric can be retrieved from the table.

[0091] In this fourth method, one hand position is fixed and the second hand's position is moved across the grid to each possible cell to calculate the probability of a hand crossing. Referring to FIGs. 22A and 22B, a grid 2210 in an illustration 2200 in FIG. 22A shows a first detection 2220 is fixed in cell (1,1) while a second detection 2230 is first placed in a next cell (1,2) of the grid 2210. Moving to an illustration 2250 in FIG. 22B, the first detection 2220 is fixed in the cell (1,1) while the second detection 2230 moves to a next cell (1,3). In this manner, the first detection 2220 remains fixed while the second detection 2230 is moved to every possible cell. A lookup table is generated with a confidence metric computed for every possible combination and stored in the memory unit for use in later determining the confidence metric. This fourth method uses the confidence metric computation from the third method. In this manner, the lookup table of confidence metric computations for each cell of the grid 2210 is computed in the pre-calculation unit 624 so that during the real time processing determination of the confidence metric is advantageously faster and computationally lightweight. An exemplary lookup table of confidence metric computations for each cell of the grid 2210 is shown in TABLE 2.

Hand 1	Hand 2	Confidence metric
(1,1)	(1,1)	Low
(1,1)	(1,2)	Low

$(1,1)$	$(1,3)$	<i>High</i>
.	.	.
$(4,3)$	$(3,7)$	<i>High</i>
.	.	.
$(10,6)$	$(10,6)$	<i>Low</i>

TABLE 2

[0092] FIG. 23 depicts identifying detected hands plotted to cells of a grid in accordance with the present embodiments. In the camera view 2300, the hands are detected. The grid 2350 corresponding to the grid 2210 in the illustrations 2200, 2250 is drawn and the cell of each hand detection (i.e., cell (4,3) and cell (3,7)) is identified. From the lookup table, it can be seen that for the cell combination (4,3) for hand one and (3,7) for hand two, the confidence metric is high. Hence, it can be considered that there is no possibility for the hands to be crossed and the leftmost hand detection can confidently be identified as a left hand and the rightmost hand detection can be confidently identified as a right hand.

[0093] Referring to FIGs. 24A and 24B, an illustration 2400 depicts a probability of a correct position occurrence of $P(A_c|p_1, p_2)$ and an illustration 2450 depicts a probability of an opposite position occurrence of $P(A_o|p_1, p_2)$. After the occurrence probability when the hand placement is correct and when it is reversed is determined for the fourth method, the confidence metric computation judges whether a combination of hand positions is confident when the difference (or ratio) is large by Equation 7 (i.e., the occurrence probability when it is correct is sufficiently large).

$$Conf = \frac{P(A_c|p_1, p_2)}{P(A_o|p_1, p_2)} \quad (7)$$

[0094] Equation 7 is one example of a computation of the confidence metric for the fourth method, alternative scoring mechanisms can also be implemented such as the mechanisms of Equations 8 and 9.

$$Conf = P(A_c|p_1, p_2) - P(A_o|p_1, p_2) \quad (8)$$

$$Conf = \ln(P(A_c|p_1, p_2)) - \ln(P(A_o|p_1, p_2)) \quad (9)$$

[0095] The occurrence probability can also be computed by using a physical model of the body as described in either the second method or the third method.

[0096] Referring to FIG. 25A, a block diagram 2500 shows the operation of the elements of the hand ID assignment unit 850 where the tracklet generator unit 852 creates tracklets, which are short sequences of object locations over time and the zone creation unit 854 creates zones of tracklets. The tracklet connector unit 856 is links tracklets based on time and spatial information to form tracks within a zone and the interpolation unit 858 fills in gaps in the tracks by interpolating between a previous detection and a next detection. The illustrations 2530, 2560 of FIGs. 25B and 25C depict a scenario showing visualizations for detected hand positions for a period of 21 frames to illustrate how tracklets are generated. The number beside each hand detection in the illustration 2530 indicates its detected frame number. Confidence metrics of each hand detection in each frame are calculated using one of the confidence metric computation methods discussed hereinabove. Hand detections in a frame having a high confidence metric indicates a leftmost detection is a left hand and a rightmost detection is a right hand.

[0097] Once the confidence metric is determined, confident and non-confident tracklets are formed based on time and spatial information. Each track in the illustration 2560 represents a tracklet. Confident tracklets consist of only hand detections with high confident tracklets and non-confident tracklets consist of at least one hand detection which is a non-confident hand detection. In the illustration 2560, in the thirteen tracklets, t5, t6, t7, t8 are non-confident tracklets and the remaining tracklets are confident tracklets.

[0098] Referring to FIG. 26A, an illustration 2600 depicts zone creation and tracklet connection in accordance with the present embodiments. A zone as shown on the illustration 2600 is created every time at least one of the hand tracklets is a non-confident tracklet and will be ended when all the hand tracklets are confident. Within each zone, connections are created between the tracklets based on time and spatial information. Candidate connections are drawn between each non-confident tracklet and other tracklets in a zone and, in order to determine which connection will be generated, a custom function is used to determine the cost of each connection. The main goal is to connect the non-confident tracklets to confident tracklets of each hand

[0099] In the constructed zone with multiple start and end tracklets, various combinations are used to calculate the shortest path cost, thereby determining the overall cost for each combination. During this process, once a shortest path is identified for a hand, the tracklets along that path are excluded from subsequent path identifications

[00100] In the end, the missing detections are interpolated based on a previous position or on an average position of previous and next detection as shown in the illustration 2650 of FIG. 26B.

[00101] Thus, it can be seen that the present embodiment provides methods and systems for determining left and right hands that is robust, reliable and computationally lightweight. Four variations of hand tracking systems and methods with confidence metric computation units in accordance with the present embodiments have been presented. These include a realtime hand tracking system and method, a realtime hand tracking system and method with a pre-calculation unit, a hand tracking system with non-realtime processing, and a combination non-realtime processing and realtime hand tracking system and method. The present embodiments also include four alternative

methods for the confidence metric computation unit 436, 628 to compute the confidence metric 516, 720. A first method is based on the difference between the x-coordinate of detected hands, the y-coordinate of detected hands, and the distance between the hands and a body position. A second method fixes a body in one position, based on the position of the wrist, elbow and shoulder joints and confidence maps are generated by a range of motion of the hands. This process is repeated for all possible body positions and confidence maps are stored in the memory unit and accessed based on a detected body position in order to compute the confidence metric. A third method involves calculating lengths of forearms, humerus and clavicles and storing the calculated lengths in the memory unit. An inverse kinematics solver is then used to estimate possible elbow and shoulder positions based on constraints like wrist position, range of rotation, movement, and lengths of forearm, humerus and/or clavicle. Based on this information, the confidence metric is computed to check for a possibility of the hands being crossed. A fourth method creates, in the pre-calculation unit 624, a grid within a region of interest and a confidence metric is computed for each possible combination of hand positions in the grid and stored in the memory unit. Any of the four methods can be utilized to create a look up table with positions and confidence metrics. Based on the position of a hand, the confidence metric can be retrieved from the lookup table.

[00102] While exemplary embodiments have been presented in the foregoing detailed description of the invention, it should be appreciated that a vast number of variations exist. It should further be appreciated that the exemplary embodiments are only examples, and are not intended to limit the scope, applicability, operation, or configuration of the invention in any way. Rather, the foregoing detailed description will provide those skilled in the art with a convenient road map for implementing an exemplary embodiment of the invention, it being understood that various changes may

be made in the function and arrangement of steps and method of operation described in the exemplary embodiment without departing from the scope of the invention as set forth in the appended claims.

CLAIMS

What is claimed is:

1. A method for left/right hand determination in a frame within a plurality of frames of a video data stream, the method comprising:

detecting one or more hands in the frame;

calculating an inter-frame distance of each of the one or more hands in the frame to a detected hand position in a previous frame of the plurality of frames;

assigning a hand ID to each of the one or more hands within the region of interest in the frame based on the calculated inter-frame distance;

computing a confidence metric based on a combination of positions of the detected hands in the frame; and

determining whether each of the one or more hands detected in the frame is a left hand or a right hand in response to the confidence metric and the assigned hand ID.

2. The method in accordance with Claim 1, wherein, when the one or more hands comprises two hands, the determining whether each of the one or more hands detected in the frame is a left hand or a right hand comprises determining a leftmost detected hand in the frame is a left hand and a rightmost detected hand in the frame is a right hand in response to the confidence metric being greater than a predetermined threshold confidence value.

3. The method in accordance with Claim 1 or Claim 2, wherein calculating the inter-frame distance comprises imputing a hand position of a missing hand based on the detected hand position in the previous frame of the plurality of frames.

4. The method in accordance with any of Claims 1 to 3, wherein detecting one or more hands in the frame comprises detecting two hands in the frame, and wherein computing the confidence metric comprises computing the confidence metric based on a function determined by a distance between x-coordinates of the two hand positions detected, a distance between y-coordinates of the two hand positions detected, and an average of the y-coordinates of the two hand positions detected.

5. The method in accordance with any of Claims 1 to 3, further comprising pre-calculating one or more confidence maps, wherein computing the confidence metric comprises computing the confidence metric based on at least one of the one or more pre-calculated confidence maps.

6. The method in accordance with Claim 5, wherein pre-calculating the one or more confidence maps comprises computing at least some of the one or more confidence maps based on ranges of motion of hands determined from positions of wrist joints, elbow joints and/or shoulder joints for every position of a person along a grid formed in a region of interest.

7. The method in accordance with Claim 5, wherein pre-calculating the one or more confidence maps comprises computing at least some of the one or more confidence maps based on an inverse kinematic estimation of possible elbow and shoulder positions determined from constraints comprising one or more of wrist position, range of rotation, movement, length of forearm, length of humerus and length of clavicle.

8. The method in accordance with Claim 5, wherein pre-calculating the one or more confidence maps comprises generating a lookup table of confidence metrics for every possible combination of hands in cells of a grid formed within a region of interest.

9. The method in accordance with any of the preceding claims, wherein determining whether each of the one or more hands detected within the region of interest in the frame is a left hand or a right hand further comprises non-realtime processing left/right hand determination in the plurality of frames of the video data stream by generating one or more tracklets of hand locations over portions of the plurality of frames, wherein each tracklet comprises a sequence of subsequently-detected hand positions all having high confident metrics or all having low confident metrics.

10. The method in accordance with Claim 9, wherein non-realtime processing left/right hand determination in the plurality of frames of the video data stream further comprises creating zones consisting of tracklets every time at least one of the tracklets is a non-confident tracklet with the tracklet ending when all the tracklets are confident tracklets.

11. The method in accordance with Claim 10, wherein non-realtime processing left/right hand determination in the plurality of frames of the video data stream further comprises connecting tracklets based on time and spatial information to form complete tracks within a zone and thereafter filling gaps in the complete tracks by

interpolating a hand position by either using a previous detection or calculating a midpoint between two hand detections.

12. A left/right hand determination system comprising:
an image capture device configured to acquire video information; and
a computing device coupled to the image capture device configured to receive the video information as a video data stream, the computing device comprising:

a hand detection unit configured to detect one or more hands in a frame of a plurality of frames in the video data stream and calculate an inter-frame distance of each of the one or more hands in the frame to a detected hand position in a previous frame of the plurality of frames, wherein the hand detection unit is further configured to assign a hand ID to each of the one or more hands in the frame based on the calculated inter-frame distance;

a confidence metric computation unit configured to compute a confidence metric based on a combination of positions of the one or more hands detected in the frame; and

a left hand/right hand assignment unit configured to determine whether each of the one or more hands detected in the frame is a left hand or a right hand in response to the confidence metric and the assigned hand ID.

13. The left/right hand determination system in accordance with Claim 12, wherein, when the one or more hands comprises two hands, the left hand/right hand assignment unit is further configured to determine a leftmost detected hand in the frame is a left hand and a rightmost detected hand in the frame is a right hand in response to the confidence metric being greater than a predetermined threshold confidence value.

14. The left/right hand determination system in accordance with Claim 12 or Claim 13, wherein the hand detection unit is further configured to calculate the inter-frame distance by imputing a position of a missing hand based on the detected hand position in the previous frame of the plurality of frames.

15. The left/right hand determination system in accordance with any of Claims 12 to 14, wherein when the hand detection unit detects two hands in the frame, the confidence metric computation unit is configured to compute the confidence metric based a function determined by a distance between x-coordinates of the two hand positions detected, a distance between y-coordinates of the two hand positions detected, and an average of the y-coordinates of the two hand positions detected.

16. The left/right hand determination system in accordance with any of Claims 12 to 14, wherein the computing device further comprises a pre-calculation unit configured to pre-calculate one or more confidence maps, and wherein the confidence metric computation unit is configured to compute the confidence metric based on at least one of the one or more pre-calculated confidence maps.

17. The left/right hand determination system in accordance with Claim 16, wherein the pre-calculation unit is configured to compute at least some of the one or more confidence maps based on ranges of motion of hands determined from positions of wrist joints, elbow joints and/or shoulder joints for every position of a person along a gris formed in a region of interest.

18. The left/right hand determination system in accordance with Claim 16, wherein the pre-calculation unit is configured to compute at least some of the one or more confidence maps based on an inverse kinematic estimation of possible elbow and shoulder positions determined from constraints comprising one or more of wrist position, range of rotation, movement, length of forearm, length of humerus and length of clavicle.

19. The left/right hand determination system in accordance with Claim 16, wherein the pre-calculation unit is configured to generate a lookup table of confidence metrics for every possible combination of hands in cells of a grid formed within a region of interest.

20. The left/right hand determination system in accordance with any of Claims 12 to 19, wherein the left hand/right hand assignment unit comprises a tracklet generator unit, and wherein the left hand/right hand assignment unit is further configured to process non-realtime left/right hand determination in the plurality of frames of the video data stream by the tracklet generator unit generating one or more tracklets of hand locations over portions of the plurality of frames, wherein each tracklet comprises a sequence of subsequently-detected hand positions all having high confident metrics or all having low confident metrics.

21. The left/right hand determination system in accordance with Claim 20, wherein the left hand/right hand assignment unit further comprises a zone creation unit which is configured to create zones consisting of tracklets every time at least one of the

tracklets is a non-confident tracklet with the tracklet ending when all the tracklets are confident tracklets.

22. The left/right hand determination system in accordance with Claim 21, wherein the left hand/right hand assignment unit further comprises a tracklet connector unit configured to connect tracklets based on time and spatial information to form complete tracks within a zone and an interpolation unit configured to thereafter fill gaps in the complete tracks by interpolating a hand position by either using a previous detection or calculating a midpoint between two hand detections.

23. A computer readable medium comprising instructions stored thereon which enable a processor to perform left/right hand determination in a frame within a plurality of frames of a video data stream in accordance with the method of any of Claims 1 to 11.