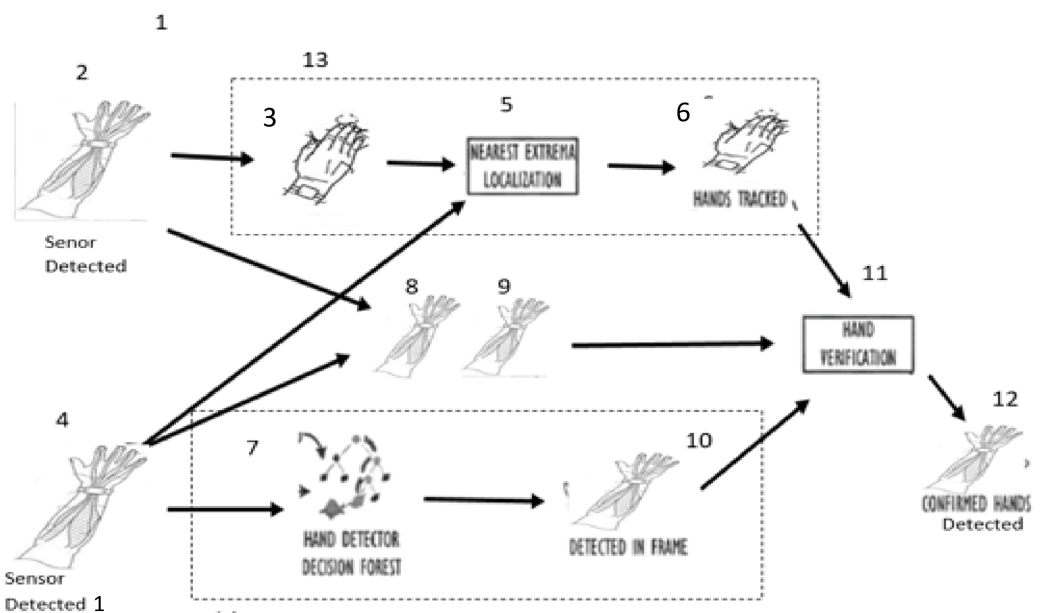
Team Slate8
Solution Design
Nathan Kebe El

Second Conceptual Design

This design description relates to a “Smart Watch” that detects a user's motion and gestures of the arm, hand, wrist, or fingers, through which the wearable Smart Watch provide American Sign Language commands to the device or to other devices which translates stored retrievable symbol representations into text or audible sound. The device is attached to, can be resting on, or in contact with parts of the users arm, wrist or hand.

One or multiple depth-sensing (3D) optical, myoelectric, mechanical contact, and or inertial sensors can detect movements of these parts of the users arm. Based on the detected movements, a user gesture can be determined. The device can translate the movement as sign language signal, match it to a stored symbol within the smartwatch and covert it to desired media.



The figure 1

A smartwatch with depth (3D), myoelectric, mechanical contact, and or inertial sensors;

a. Computer code programs stored in memory (three dimensional sensors, sensor tracking, sequencing, pattern learning algorithms, ASL patterns or descriptors, etc.)
b. and one or more processing units operatively coupled to the memory and configured to execute instructions in the computer program code that cause the one or more processing units to:
    i. receive three dimensional sensor coordinates of an area containing at least one human finger from the smartwatch sensors,
    ii. the 3D sensor depth map comprising a matrix of coordinates, each coordinate having a depth value;
c. extract, from the depth coordinates, descriptors based on the depth values of the coordinates in a plurality of patches distributed in respective positions over at least one human finger;
d. match the extracted descriptors to previously-stored descriptors using a background-invariant decision forest;
e. estimate a position of at least one of the human fingers based, at least in part, on stored information associated with the matched descriptors;
f. track the position of at least one human fingers bi-directionally along a z-axis of the area.
g. The descriptors use a constant value for the depth values of background coordinates.

The figure 1, illustrates an exemplary process (1) for performing three dimensional sensor hand or finger sensor tracking using sensor depth sequences, in accordance with some embodiments. First, the depth coordinates for a previous map (Sensor detected) is analyzed at Step (2). The depth coordinates data for the previous map may simultaneously be passed to sensor tracking process (13) as well as motion detection process (8) and background modeling process (9)(for further verification. Within sensor tracking process (13), the hand detected in the Sensor are obtained (3). Next, the hand or fingers from Sensor detected and Map (Sensor detected 1) (i.e., the “current map”) may be subjected to a nearest extrema localization constraint (5), which effectively searches the area surrounding the location of the hand in the previous map and constrains any hand candidate that is detected in the map to be located at the nearest extrema of a foreground body object, otherwise it can be discarded. The hand or fingers that pass the nearest extrema localization constraint at step (5) are the confirmed set of hand or fingers that will be tracked (6) from Sensor detected.

Like the previous map (Sensor detected), the information from the current map (sensor detected) may be passed to motion detection process (8) and background modeling process (9). The result of processing the previous and current map using motion detection process (8) and background modeling process (9) is that the portions of the maps where a valid hand may be tracked are limited to only “moving coordinate areas” within the map, as well as coordinates that are deemed “foreground coordinates” within the map. These two constraints come from the insight that true positive hand or fingers in received coordinates are almost always both in the foreground, as well as moving. (Note: the motion detection process (8) and background modeling process (9) may “look back” a set number of maps, e.g., a few seconds' worth of maps, to determine whether there is a high enough probability of movement in a particular region of the coordinates to deem it a “motion region” or a “background region.”)

Finally, each map, as it becomes the “current map,” is subjected to the hand detection process (14), which may involve the background-invariant hand detector decision forest (7) discussed in greater detail above, resulting in a set of candidate hand or fingers (10) detected in the “current map,” (Sensor detected 1). Note that there may be some new hand or fingers in “Sensor detected 1” that were not present in “Sensor detected”, or some hand or fingers that were present in “Sensor detected” that are not present in “Sensor detected 1”. This output of current map hand detection process (14) is then passed to hand verification stage (11), along with: 1.) the output of motion detection process (8) and background modeling process (9) that limit the potential parts of the map where valid hand or fingers can appear to the moving portions of the foreground; and 2.) the output of the hand tracker (13) from the previous map.

At hand verification step (11), the detected (10) and tracked (6) hand candidates are again verified to make sure that they are likely to be hand or fingers. This process again leverages the fact that the hand is most often attached to the body with “single directional connectivity,” i.e., located at the end of an arm that is connected to the user's body via only a single connection point. This “single directional connectivity” check may be implemented by drawing a circle around the detected hand and checking the intersections with the foreground body mass of the user whose hand has been detected. If the detected hand is indeed a real hand, there will only be an intersection with the foreground body mass in one direction.

Additional verification steps would include ruling out candidate hand or fingers located in the background of the scene or in regions of the scene where there has not been any movement over a predetermined amount of time. Finally, hand or fingers may be verified by sensor tracking their IDs from map to map and removing those hand or fingers in the current map that show unusual movement characteristics. For example, if there are two hand or fingers in the map for many consecutive maps, and then there are suddenly six hand or fingers in the current map, there is a high likelihood that four additional hand or fingers in the current map may be false positives. Likewise, if Hand #1 has been on the left side of an coordinates for many consecutive maps and Hand #2 has been on the right side of an coordinates for many consecutive maps, it is unlikely that, in the current map, either Hand #1 or Hand #2 would suddenly move all the way across to the other side of the coordinates (i.e., move more than the distance by which a human subject could typically move their hand or fingers in the time it took the coordinates sensor to capture the successive maps). Finally, those hand candidates that pass the hand verification step (11) are output as confirmed hand or fingers in the current map, Sensor detected, (12).

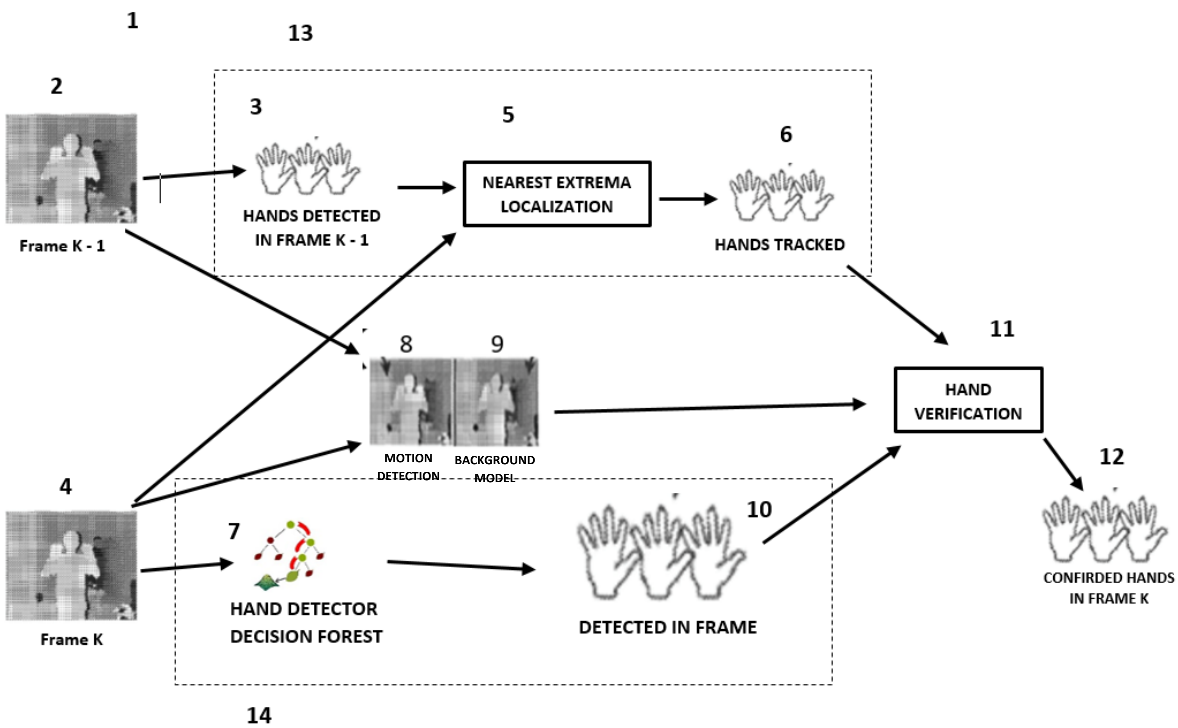
A random decision tree from a random decision forest, in accordance with some embodiments, e.g., the random decision forest that is created during the hand detector training process or the random decision forest (7) that is used during the hand detection process once it has been trained. Each decision tree comprises a root node a plurality of internal nodes, called split nodes, and a plurality of leaf nodes.

Final Conceptual Design Selection: Top Level Design

Future smartphones software and cameras can support sign language recognition.

In the study of Human-computer interaction (HCI) or people interacting with computers, understanding how people interact with computers is an important mission. The capability to recognize human gestures, in specific, hand and head movements in relation to HCI, is a very vital part in grasping the objectives and needs of the user in an extensive range of applications. In this innovation, a unique system and technique aimed at 3-D hand tracking using depth sequences is defined. The hand tracking method defined here covers: 1.) a strong hand detector that is never changing in relation to scene background changes; 2.) a two - directional tracking procedure that stops detected hands drifting nearer or closer towards the front of the scene (i.e., advancing along the z-axis of the scene); and 3.) Numerous hand validated heuristics.

Three-Dimensional Hand Tracking With Depth Sequences, describes how devices can locate and follow the positions of hands over 3-D space in video streams, akin to face-tracking technology already used in smartphones camera apps. Using this technology for the ASL is very possible for detecting hand gestures.



Description:

A depth-sensing camera;

a. Computer code programs stored in memory (three dimensional, tracking, sequencing, pattern learning algorithm, ASL patterns or descriptors, etc.)
b. and one or more processing units operatively coupled to the memory and configured to execute instructions in the computer program code that cause the one or more processing units to:
    i. receive a depth map of a scene containing at least one human hand from the depth-sensing camera,
    ii. the depth map comprising a matrix of pixels, each pixel having a depth value;
c. extract, from the depth map, descriptors based on the depth values of the pixels in a plurality of patches distributed in respective positions over at least one human hand;
d. match the extracted descriptors to previously-stored descriptors using a background-invariant decision forest;
e. estimate a position of at least one of the human hand based, at least in part, on stored information associated with the matched descriptors;
f. and track the position of at least one human hand bi-directionally along a z-axis of the scene.
g. The descriptors use a constant value for the depth values of background pixels.

The figure 1, illustrates an exemplary process (1) for performing three dimensional hand tracking using depth sequences, in accordance with some embodiments. First, the depth image for a previous frame (Frame K−1) is analyzed at Step (2). The depth image data for the previous frame may simultaneously be passed to tracking process (13) as well as motion detection process (8) and background modeling process (9)(for further verification. Within tracking process (13), the set of hands detected in the Frame K−1 are obtained (3). Next, the hands from Frame K−1 and Frame K (i.e., the “current frame”) may be subjected to a nearest extrema localization constraint (5), which effectively searches the area surrounding the location of the hand in the previous frame and constrains any hand candidate that is detected in the frame to be located at the nearest extrema of a foreground body object, otherwise it can be discarded. The hands that pass the nearest extrema localization constraint at step (5) are the confirmed set of hands that will be tracked (6) from Frame K−1.

Like the previous frame (Frame K−1), the information from the current frame (Frame K) may be passed to motion detection process (8) and background modeling process (9). The result of processing the previous and current frame using motion detection process (8) and background modeling process (9) is that the portions of the frames where a valid hand may be tracked are limited to only “moving pixel areas” within the frame, as well as pixels that are deemed “foreground pixels” within the frame. These two constraints come from the insight that true positive hands in received images are almost always both in the foreground, as well as moving. (Note: the motion detection process (8) and background modeling process (9) may “look back” a set number of frames, e.g., a few seconds' worth of frames, to determine whether there is a high enough probability of movement in a particular region of the image to deem it a “motion region” or a “background region.”)

Finally, each frame, as it becomes the “current frame,” is subjected to the hand detection process 14, which may involve the background-invariant hand detector decision forest (7) discussed in greater detail below, resulting in a set of candidate hands (10) detected in the “current frame,” Frame K. Note that there may be some new hands in Frame K that were not present in Frame K−1, or some hands that were present in Frame K−1 that are not present in Frame K. This output of current frame hand detection process (14) is then passed to hand verification stage (11), along with: 1.) the output of motion detection process (8) and background modeling process (9) that limit the potential parts of the frame where valid hands can appear to the moving portions of the foreground; and 2.) the output of the hand tracker (13) from the previous frame.

At hand verification step (11), the detected (10) and tracked (6) hand candidates are again verified to make sure that they are likely to be hands. This process again leverages the fact that the hand is most often attached to the body with “single directional connectivity,” i.e., located at the end of an arm that is connected to the user's body via only a single connection point. This “single directional connectivity” check may be implemented by drawing a circle around the detected hand and checking the intersections with the foreground body mass of the user whose hand has been detected. If the detected hand is indeed a real hand, there will only be an intersection with the foreground body mass in one direction.

Additional verification steps would include ruling out candidate hands located in the background of the scene or in regions of the scene where there has not been any movement over a predetermined amount of time. Finally, hands may be verified by tracking their IDs from frame to frame and removing those hands in the current frame that show unusual movement characteristics. For example, if there are two hands in the frame for many consecutive frames, and then there are suddenly six hands in the current frame, there is a high likelihood that four additional hands in the current frame may be false positives. Likewise, if Hand #1 has been on the left side of an image for many consecutive frames and Hand #2 has been on the right side of an image for many consecutive frames, it is unlikely that, in the current frame, either Hand #1 or Hand #2 would suddenly move all the way across to the other side of the image (i.e., move more than the distance by which a human subject could typically move their hands in the time it took the image sensor to capture the successive frames). Finally, those hand candidates that pass the hand verification step (11) are output as confirmed hands in the current frame, Frame K, (12).
A random decision tree from a random decision forest, in accordance with some embodiments, e.g., the random decision forest that is created during the hand detector training process or the random decision forest (7) that is used during the hand detection process once it has been trained. Each decision tree comprises a root node a plurality of internal nodes, called split nodes, and a plurality of leaf nodes.