August 2020

Hand Input Gesture Recognition Using ML-based Pose Estimation and Optical Flow Processing

Anonymous

Hand Input Gesture Recognition Using ML-based Pose Estimation and Optical Flow Processing

ABSTRACT

Current techniques for detection of hand pose have high computational cost, and have accuracy and speed limitations which make them less than satisfactory for use in some high precision and time-sensitive applications. This disclosure describes the use of a combination of machine learning based pose estimation and optical flow processing techniques for hand gesture recognition, e.g., in AR and VR applications. Machine learning (ML) based pose estimation is utilized to identify hand pose and identify areas of interest that are processed further using optical flow processing. The use of optical flow processing enables higher precision in the detection of small movements as well as good low light performance. The use of ML based pose estimation reduces the solution space by reducing the size of an image that is subjected to optical flow processing. Such reduction of the solution space reduces the computational load and time required for optical flow processing.

KEYWORDS

- Hand gesture
- Gesture recognition
- Optical flow
- Pose estimation
- Virtual Reality (VR)
- Augmented Reality (AR)
- Mixed Reality
- Hybrid Reality
- Artificial Reality

BACKGROUND

Augmented Reality (AR) and Virtual Reality (VR) wearables such as glasses commonly use hand gestures as a mode of user input to control and/or otherwise interact with the AR/VR glasses. For example, users can make gestures, e.g., hand movements, pinching, scrolling, swiping, etc. for performing actions such as selecting objects, moving objects, activating applications, etc. User experience can be greatly enhanced by accurate and rapid recognition (interpretation) of user hand gestures. Currently, hand pose estimation techniques are utilized for gesture recognition. These techniques have accuracy and speed limitations which can make them less than satisfactory for use in some high precision and time-sensitive applications such as gaming, discrete inputs, text entry, etc.

DESCRIPTION

This disclosure describes techniques for gesture recognition in AR and VR applications. Per techniques of this disclosure, machine learning based hand pose estimation is combined with optical flow processing for fast and accurate gesture recognition. For example, optical flow processing can be performed for a specific region of the captured image, e.g., the fingertip area.

Pose estimation is utilized to infer the pose and/or motion of a person or object in image(s) captured by the wearable device. Pose estimation typically includes identifying, locating, and tracking key points located on an object or person of interest. For example, to perform pose recognition for humans, the key points can include major joints such an elbow or knee.

Optical flow processing is the processing of data, e.g., sensor data to infer the motion of the observer and objects in a scene represented by the data and details of the structure of objects and the environment in the scene. It offers greater precision when compared to other techniques such as pose estimation. However, optical flow processing has higher computational requirements.

Per techniques of this disclosure, machine learning based pose estimation is utilized to identify hand pose and identify areas of interest. The identified areas are further analyzed using optical flow processing. The combination of the two techniques to analyze a scene enables the extraction of discrete user movement information with a higher degree of accuracy than with just ML based pose estimation. The combination also has a lower processing cost than applying optical flow processing to the entire scene.

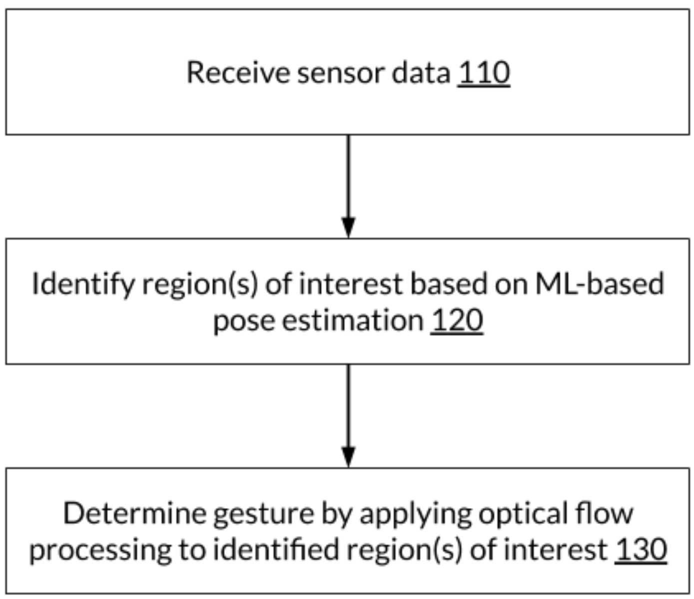

Fig. 1: Gesture recognition using ML based pose estimation and optical flow processing

Fig. 1 illustrates an example process for gesture recognition, per techniques of this disclosure. Sensor data is received (110) from one or more sensors of a user device. For example, the sensor data can include image(s) from one or more cameras (e.g., RGB cameras, infrared cameras, etc.) located on an AR/VR headset. In another example (in addition to or separate from the previous example), the sensor data can include motion data and biological data, such neuro-muscular data.

Pose estimation is performed on the received sensor data to identify region(s) of interest (120). Pose estimation can include the utilization of a machine learning model trained using hand gesture training data. For example, pose estimation can be utilized to identify data segment(s) or regions in the sensor data where a gesture is most likely to be present.

Based on the identified region(s) of interest, the received sensor data is selectively analyzed using optical flow processing for precise determination of a user gesture (130). The determined gesture can be provided to any application which can then perform the corresponding operation, e.g., perform control tasks with reference to an object, etc.

The use of optical flow processing enables improved accuracy in timing and higher precision in the detection of small movements as well as improved low light performance in comparison to just ML based pose estimation. For example, optical flow processing can improve the accuracy of recognition of fine hand movements and discrete activations based on those movements, e.g., tapping, pinching, etc. Additionally, accuracy of optical flow processing is improved since use of ML based pose estimation isolates observer movement (user head movement) from hand movement.

By reducing the size of an image to be processed by optical flow processing, ML based pose estimation provides a reduction of the solution space for the optical flow processing. The reduction of the solution space corresponding reduces the computational time and computational load (e.g., CPU/GPU power) utilized for optical flow processing. Further, identifying an accurate range of observer-to-hands relationship in this manner can reduce the need to re-process hands through the pose estimation machine learning model, which can also save computational requirements (battery, processing power, time, etc.).

While the foregoing discussion refers to a wearable device such as an AR/VR headset, the described techniques can also be utilized for providing gesture input to a console controller, e.g., a gaming console or other device that utilizes cameras to detect user gestures.

CONCLUSION

Current techniques for detection of hand pose have high computational cost, and have accuracy and speed limitations which make them less than satisfactory for use in some high precision and time-sensitive applications. This disclosure describes the use of a combination of ML based pose estimation and optical flow processing techniques for hand gesture recognition, e.g., in AR and VR applications. Machine learning (ML) based pose estimation is utilized to identify hand pose and identify areas of interest that are processed further using optical flow processing. The use of optical flow processing enables higher precision in the detection of small movements as well as good low light performance. The use of ML based pose estimation reduces the solution space by reducing the size of an image that is subjected to optical flow processing. Such reduction of the solution space reduces the computational load and time required for optical flow processing.