EXPLORING MOTION CAPTURE ALGORITHMS IN COMPUTER VISION USING INTEL DEPTH CAMERA

Ollaberganova Muyassar
Graduate student Urganch branch of TUIT olloberganovamuyassar@gmail.com

Abstract: The analysis of existing approaches to tracking the human body revealed the presence of problems when capturing movements in a three-dimensional coordinate system. The promise of motion capture systems based on computer vision is noted. Existing research on markerless motion capture systems only considers positioning in 2D space. Therefore, the goal of the study was to improve the accuracy of determining the coordinates of the human body in three-dimensional coordinates by developing a motion capture method based on computer vision and triangulation algorithms.

Significant progress has now been made in the field of computer vision. Technologies have been developed that allow solving the problems of detecting objects, determining their state, geometric assessment of the space depicted in the frame, and many others. Thanks to this, computer vision has become widespread in various fields of human activity, from healthcare and education to the entertainment sector. A fairly promising direction is the use of computer vision technologies for three-dimensional reconstruction and positioning of various objects, including people. There are quite a large number of systems for determining the absolute position of a person in space, which can be divided into the following categories:

 systems that use inertial sensors and make it possible to determine the magnitude of their movement, as well as changes in angles between them, which involves the use of gyroscopes and accelerometers [1]. A well-known representative of this category is Intel Depth [2], which includes up to 32 inertial sensors;

 laser positional tracking systems, based on the use of base stations installed on opposite sides of the room and emitting infrared rays, which make it possible to accurately determine the position and orientation of sensors in space. An example of such systems are Intel Depth virtual reality kits from HTC [3], which have an error of up to 0.1 mm;

 systems using magnetic sensors [4], based on the use of a magnetic field to capture human movement, which involve the presence of wearable sensors on the user's body. Intel Depth falls into this category.

- portable electromagnetic motion tracking system, considered one of the fastest (sampling frequency 240 Hz);
 optical systems based on markers - determine the position of objects using markers using a set of cameras. An example is Intel Depth, which has a fairly low error: the average absolute marker tracking errors are 0.15 mm in static tests and 0.2 mm (with corresponding angular errors of 0.3°) in dynamic tests [5];
 markerless optical systems based on the use of computer vision and machine learning. Examples of such technologies are OpenPose, MediaPipe, Intel Depth. With their help, human movements can be tracked with an accuracy of up to 30 mm [6].

Keywords: motion capture, virtual reality, triangulation, computer vision, machine learning.

Analyzing the listed categories of motion capture systems, we can conclude that most solutions used to recognize human actions and movements involve the presence of various wearable devices, such as sensors or gloves. The bulk of these devices are cumbersome due to the large number of sensors and the need for a wired connection. Some such systems have high accuracy, but cannot be used due to their size or the presence of electromagnetic interference [7]. Inertial systems have a number of problems associated with error accumulation, which limits their use to relative positioning in space only.

Materials and methods. A motion capture method is presented, including calibration of multiple cameras and formalization of procedures for detecting a person in a frame using a convolutional neural network. Based on the skeletal points obtained from the neural network, a three-dimensional reconstruction of a model of the human body is carried out using various triangulation algorithms.

Thus, as part of the second stage, a set of 33 points is formed for each i-th camera:

$$\left\{x_{ij} = \left\langle u_{ij}, v_{ij} \right\rangle \mid j \in \{1,2,\ldots,33\}, i \in \{1,2,\ldots,K\}\right\},$$

where ij u is the coordinate of the j-th point along the X axis on the i-th image; ij v is the coordinate of the j-th point along the Y axis on the i-th image; K is the total number of cameras and images.

Research results. Experimental studies were conducted to compare four triangulation algorithms: direct linear transfer, linear least squares, L2 triangulation and polynomial methods. The optimal triangulation algorithm (polynomial) has been determined, providing an error of no more than 2.5 pixels or 1.67 centimeters.

Discussion and conclusion. The shortcomings of existing motion capture systems have been identified. The proposed method aims to improve the accuracy of motion capture in 3D coordinates using computer vision. The results obtained are integrated into software for positioning the human body in three-dimensional coordinates for remote monitoring, use in virtual simulators and motion capture systems.

BIBLIOGRAPHY

1. Lind C.M., Abtahi F., Forsman M. Wearable Motion Capture Devices for the Prevention of Work-Related Musculoskeletal Disorders in Ergonomics – An Overview of Current Applications, Challenges, and Future Opportunities. Sensors. 2023;23(9):4259.
2. Sers R., Forrester S., Moss E., Ward S., Ma J., Zecca M. Validity of the Perception Neuron Inertial Motion Capture System for Upper Body Motion Analysis. Measurement. 2020;149:107024.
3. Bauer P., Lienhart W., Jost S. Accuracy Investigation of the Pose Determination of a VR System. Sensors. 2021;21(5):1622.
4. Irshad M.T., Nisar M.A., Gouverneur P., Rapp M., Grzegorzek M. AI Approaches towards Prechtl's Assessment of General Movements: A Systematic Literature Review. Sensors. 2020;20(18):5321.
5. Merriaux P., Dupuis Y., Boutteau R., Vasseur P., Savatier X. A Study of Vicon System Positioning Performance. Sensors. 2017;17(7):1591.