In recent years, many papers have been conducted to estimate pose from one image using deep learning. In this paper, we apply pose estimation using deep learning and support posture training of trunk training using camera images. Since the trunk plays an important role with regard to motor connection and balance between limbs, its importance has also been drawing attention for the general public. And trunk training that trains this trunk can generally be done by an individual, and it is important to maintain a correct posture in order to maximize its effect. However, it is difficult for individuals to grasp posture during training. Therefore, there is a method of grasping posture through images taken with mirrors and cameras, but it is also difficult to determine whether the grasped posture is correct or not. From this, it can be said that attitude assistance in trunk training is effective.

There are papers to support pose estimation of training, feedback to the user, prompt training with correct motion and posture, but since it requires a depth camera and multiple cameras to estimate their posture, There will be a cost. Also, since the visual feedback to the user is a 2D bone image showing only skeleton information in one viewpoint, it is difficult to grasp the posture. Therefore, in this paper, we only use a single RGB camera to estimate pose from the camera image, generate a 3D model, and give visual feedback to the user in two viewpoints, thereby supporting posture support of trunk training. Furthermore, in order to make it easy to grasp the attitude, feedback is provided from the viewpoint the user wishes to see. By supporting from the camera image, it is possible to acquire the attitude of training even in reference books and images such as the Internet, so it can be easily applied to various kinds of events.

This paper aims to propose a posture support system using only a single RGB camera. Further Aiming at a system that can easily correct posture by making visual feedback that the user can easily grasp posture. By using this system, you can train with individuals, in the right posture without going to the gym where the trainer is located. The contents of this paper can be divided into two types: generation of 3D model from camera image and posture support.

In generating a 3D model from a camera image, normalize the image using a bounding box before pose estimation is performed on images. The bounding box is a rectangle including a person. Pose estimation is performed using Human Mesh Recovery (HMR) proposed by Kanazawa et al. On the normalized image, and a 3D model is generated using the estimated parameters.

As the flow of the posture support system as a whole, the process of generating a 3D model using the training image before training and the camera image under training is similar, but the subsequent processing is different. Before training, save viewpoint and 3D model, and generate visual feedback during training. The user inputs an image of the training to be performed before training. The system generates a 3D model from the input image. Then, the user sets a viewpoint in which the orientation of the generated 3D model can be easily understood. The viewpoint and the generated 3D model are saved for use in processing during training.

During training, the training image of the user is acquired from the Web camera, and similarly to the processing before the training, the 3D model is generated. Then, as a visual feedback to the user, the 3D model of the generated current posture and the 3D model of the pre-stored target posture are superimposed and displayed. Also, in order to make it easy to understand the difference between the target posture and the current posture, markers are displayed at ten parts (both hands and elbows, shoulder, knee, ankle) of the body, and according to the error from the target posture, Are changed in four stages to urge the correction of the posture. The user brings the posture closer to the target posture based on these presented information. With the above flow, this system supports posture support to the user and encourages training in the correct posture.

Evaluation experiments and questionnaire surveys are conducted to verify the usefulness of the method

proposed in this paper. Implemented in this system with a method (method 1) for presenting 2D bones used as feedback information in related research with one viewpoint and a method (method 2) for presenting the 3D model proposed in this paper from two viewpoints Then compare. Root Mean Squared Error (RMSE) was used for comparison of experimental results. In order to investigate the influence on posture correction, the average value of the ratio obtained by subtracting the initial value from the minimum value of RMSE is divided by the initial value, and as a result, method 2 is about 7% larger than method 1. From this fact, we estimate that method 2 has a greater influence on posture correction. It is also suggested that there is no big difference in the average value of the time to reach the minimum value.

As a result of the questionnaire, we assume that the method proposed in this paper is effective because everyone of the subjects answers that it is easy to grasp the attitude of the 3D model from two viewpoints. Also, since the average value of responses to the question of whether it is possible to change the viewpoint or to improve ease of grasping posture by performing feedback using a 3D model was high, they are effective methods Think.

In conclusion, we implemented a method for presenting 2D bones used as feedback information in related research from one viewpoint and a method for presenting the 3D model proposed in this research from two viewpoints in this system and evaluate comparison was carried out. Using the Root Mean Square Error (RMSE), we compare the 3D model of the target posture with the estimated 3D model of the posture, but no significant difference was found. However, the questionnaire conducted at the same time is considered to have the effectiveness of the proposed method.

The problem with this system is the slow update speed of visual feedback. Currently it is updated once every 2 seconds. Since feedback does not return immediately in actual use, the error from the target posture at that time is unknown and corrected, so that there is a possibility that it will go away from the target posture. Therefore, high-speed processing is desired. Also, with the training image as input, it is hard to say that the generated 3D model of the target attitude is consistent with the input image. In order to provide better support, it is also desirable to improve the estimation accuracy. When using this system, there is also a problem that the display must always be seen. We think that this problem can be solved by using HMD (Head Mounted Display).