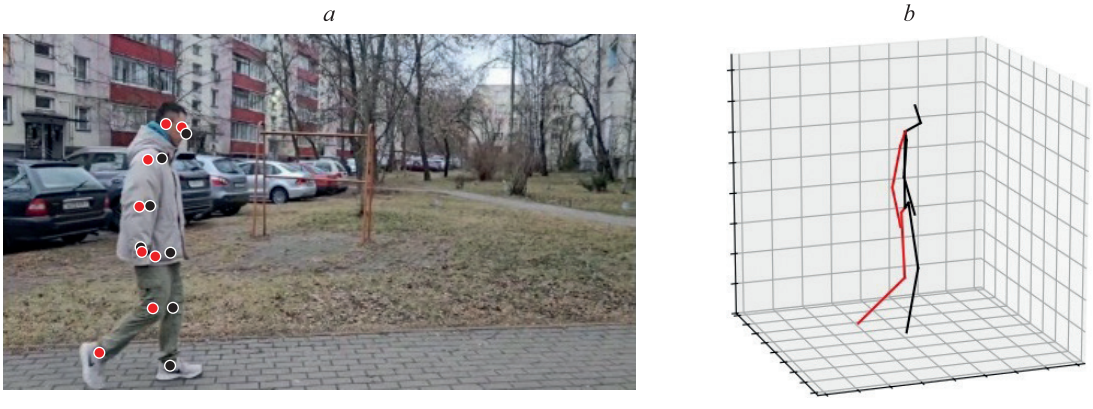

Fig. 4. 3D pose estimation using the MTSA-former model:
a – input 2D human keypoint detection;
b – reconstructed 3D human skeleton predicted by the model