BRAIN. Broad Research in Artificial Intelligence and Neuroscience
e-ISSN: 2067-3957 | p-ISSN: 2068-0473
Covered in: Web of Science (ESCI); EBSCO; JERIH PLUS (hkdir.no); IndexCopernicus; Google Scholar; SHERPA/RoMEO; ArticleReach Direct; WorldCat; CrossRef; Peeref; Bridge of Knowledge (mostwiedzy.pl); abcdindex.com; Editage; Ingenta Connect Publication; OALib; scite.ai; Scholar9; Scientific and Technical Information Portal; FID Move; ADVANCED SCIENCES INDEX (European Science Evaluation Center, neredataltics.org); ivySCI; exaly.com; Journal Selector Tool (letpub.com); Citefactor.org; fatcat!; ZDB catalogue; Catalogue SUDOC (abes.fr); OpenAlex; Wikidata; The ISSN Portal; Socolar; KVK-Volltitel (kit.edu)
2025, Volume 16, Special Issue 1 (April 2025), pages: 117-126.
Special Issue 1: *Neuroscience, Artificial Intelligence, and Innovation in Education*
Submitted: December 22nd, 2024 | *Accepted for publication:* February 17th, 2025

Research on Motion Capture Technologies and AI Video Synthesis for Creating Digital Bulgarian Folk Choreographies

Donika Valcheva
Department of Mathematics and Informatics, St. Cyril and St. Methodius University, Veliko Tarnovo, Bulgaria.
d.valcheva@ts.uni-vt.bg
https://orcid.org/0000-0003-2107-0190

Teodor Kalushkov
Department of Mathematics and Informatics, St. Cyril and St. Methodius University, Veliko Tarnovo, Bulgaria.
t.kalushkov@ts.uni-vt.bg
https://orcid.org/0000-0002-5748-6238

Georgi Shipkovenski
Department of Mathematics and Informatics, St. Cyril and St. Methodius University, Veliko Tarnovo, Bulgaria.
g.shipkovenski@ts.uni-vt.bg
https://orcid.org/0000-0002-7627-0635

Abstract: *The article is focused on the following three scientific and applied activities: research into motion capture technologies, analysis of applications for AI video synthesis, and developing a methodology for creating digital choreographies of Bulgarian folk dances. The possibilities for automation through AI and MoCap were assessed, including extraction, adaptation, and synchronisation of choreographic movements with virtual avatars using retargeting technologies. When analysing the leading AI-based applications for markerless MoCap, it was found that they differ in the detail of motion capture, complexity of use, hardware and software requirements. Methods for 3D reconstruction of people from images, significantly facilitates the process of 3D modelling but require specific additional animation settings that improve their work, especially when used to create realistic movements in a digital environment. Automated rigging and animation tools are extremely useful for rapid prototyping, although in some cases additional processing is required to achieve realistic movements. Effective software solutions were identified, a comparative analysis of different AI tools was made and recommendations for the selection of technologies were proposed, based on their purpose, degree of complexity, need for additional software/hardware, price and compatibility with Motion Capture. Based on this research, a step-by-step methodology for creating digital choreographies of Bulgarian folk dances was developed. It integrates MoCap, AI-based tools and automated animation techniques.*

Keywords: *motion capture; synchronisation; AI video synthesis; 3D scene; digital choreographies.*

How to cite: Valcheva, D., Kalushkov, T., & Shipkovenski, G. (2025). Research on motion capture technologies and AI video synthesis for creating digital Bulgarian folk choreographies. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, *16*(Sup1), 117-126. https://doi.org/10.70594/brain/16.S1/10

©2025 Published by EduSoft Publishing. This is an open-access article under the CC BY-NC-ND license

1. Introduction

The rapid development of digital technologies leads to a certain degree of digital transformation in every area of human life. Technologies and artificial intelligence (AI) were found long before their place in the art of dance. They are integrated very successfully in the form of video projections, holograms, 3D effects, computer-controlled lighting and sound, etc. Artificial Intelligence makes creating video content extremely easy with the wide range of AI video generators available.

The purpose of the report is expressed in the following three scientific and applied activities: research of technologies for capturing the movements of the human body, research of applications for AI video synthesis, and development of a methodology to create digital choreographies of Bulgarian folk dances.

2. Research On Technologies For Capturing The Movements Of The Human Body

Researching and popularising the Bulgarian folklore heritage are important factors for the preservation of our national identity. Motion capture is the process of recording the movement of people and objects. In cinematography and video game production, the process captures the movement of human actors and uses the information to animate 2D and 3D digital character models, providing a significant level of realism. Motion Capture is also used in other areas, such as the analysis of movements during physical activities related to sports or rehabilitation activities (Salisu et al., 2023).

There are various motion capture technologies, which are most often classified by the method of motion detection and the hardware used (Figure 1).

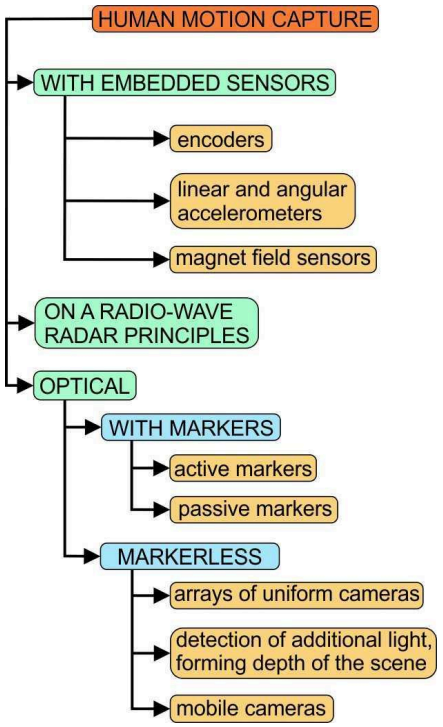


Figure 1. Classification of human motion detection methods

1) A large part of the technological solutions for Motion Capture are based on sensors attached to moving bodies.

These are three-axis linear and angular acceleration sensors or encoders. To be attached to the human body, they are often incorporated into specialised suits, socks, gloves, bracelets, etc.

In addition to the inconvenience they create, the difficulties with these systems are also related to the data transmission from all the sensors in real-time to the node for recording and centralised management of the data. These systems require prior calibration (resetting) of all sensors.

2) In some implementations, magnetic field-detecting sensors are used instead of acceleration sensors. It is possible to use the Earth's magnetic field or a dedicated source (Rahul, 2018). The weakness of these solutions is related to the limitations regarding the environment to metallic objects and magnetic disturbing sources in the vicinity of the surveyed objects. Again, the initial calibration of the sensors is used. Some solutions combine acceleration and magnetic sensors in order to achieve a higher accuracy of motion recognition.

3) Another part of Motion Capture systems uses passive marker elements that attach to the surface of moving objects. It is possible to position them on costumes, which, in addition to the individual marker points, also have lines drawn between them to aid the analysis of the skeleton structures. To detect the movement, cameras are used that record the image coming through the lens to the matrix.

It then filters out the light reflected by the markers and possibly the skeleton connections between the markers with a specific wavelength. Next is the creation of the virtual models. These solutions usually use a large number of cameras located in different positions. Calibration is also necessary, but unlike sensor systems, in this case, it aims to synchronise the position of moving objects relative to cameras located at different angles.

4) Active marker elements, unlike passive ones, emit light of a certain wavelength. Thanks to hardware filters or the use of computer vision software techniques, marker movements are detected from the moving object. An additional disadvantage of these methods is related to the need to power the active markers. It is implemented either by small batteries located in each marker or by wiring along the body of the moving object.

5) An alternative to fixed markers is the use of structured light projectors. In most cases, it is infrared and consists of spots of different intensities that are imposed on the analysed objects. In addition to capturing the scene in full colour, structural infrared reflections are also extracted from it by filtering. They form the depth of the scene, thus eliminating the need to use multiple cameras. Calibration in this setup is expressed in synchronising the position of moving objects relative to the two-dimensional image and the depth image based on the infrared channel. The method is significantly cheaper and easier to implement, but from the point of view of the spatial aspect, it has serious disadvantages compared to the methods using multiple cameras. The reason is that the analysis of moving objects is based on a single observation position.

6) Instead of using light, some motion detection systems rely on acoustic methods. The waves used are most often from the ultrasonic spectrum. A radar principle is used to detect the propagation time of the wave. Systems based on these principles are of limited use due to their serious drawbacks related to interference that occurs as a result of reflected waves and, in certain cases, from external sources. This imposes restrictions on their use both in open spaces and indoors. Recognising fin movements in group scenes also proves to be a serious challenge here.

7) Markerless 3D motion detection based only on cameras located from different positions of the studied object is also used in many cases. Its advantages are related to the simplified hardware implementation and the freedom of movement of the studied objects. However, this choice relies on highly sophisticated software that uses computer vision techniques and is distinguished by a serious computational burden. It imposes high demands on the computers used, both in terms of central computing resources and graphics. The accuracy and resolution of the achieved results mainly depend on the hardware used and the selection of the algorithms implemented in the software device. In this case, preliminary calibration is used, in which the person whose movements are being monitored stands in a certain pose, in a certain position. In addition to the use of fixed cameras, some variations of this family of methods use one or more mobile cameras, most often attached to drones. After detecting movements, to recreate them in the future, it is necessary to integrate them into objects created and studied so far in the virtual environment. Specifically, in humans and

animals, skeletal structures are most often used. Depending on the level of detail that is desired, the skeleton can be constructed with a different number of joints.

It is also possible to integrate muscle models, but this greatly complicates the task. With a model with a large number of joints and a complex muscle model, the processing needs to go beyond the scope of desktop computers and move to artificial intelligence systems. This migration, however, greatly increases the level of realism when playing scenes generated by the three-dimensional environment.

The use of skeleton models is necessary due to the limitations of the motor possibilities in the human spine and joints. Similarly, muscle patterns are dictated by the uneven distribution of muscle mass in the human body and the disproportionate elasticity of tissues in its different areas. In the absence of the specified models, overexposure to moving objects can be observed in the reconstructed scenes, which manifests itself most often as unnatural positions and movements (for example, turning a human head 360º relative to the torso).

Having in mind the set goals of the research, when choosing a specific Motion Capture technology, the following criteria should be taken into account:

- The detected movements should be transferred into 3D scenes so that the choreography can be replayed subsequently with a large number of participants.
- The detected movements must not be analysed independently but must be classified according to the movements in the dance.
- The detection resolution should be at the level of limbs and fingers, without the need to examine the face.
- It is possible that after the capture of the movements, it is necessary to filter them (Uzunova, Chotrov, & Maleshkov, 2016) before submitting them for further processing.
- In view of the practical importance of the research and the limited budget of the small dance halls, it is necessary to take into account the financial dimensions of the possible solutions.

3. Exploring AI Video Synthesis Applications

One of the possibilities of artificial intelligence is the generation and synthesis of video. Generation refers to the creation of new video content based on inputs such as text descriptions, audio recordings, body positioning, scene context, and more. Synthesis, on the other hand, involves combining and transforming elements of pre-existing video material to create a new version of the same video with modifications or corrections.

This possibility is becoming a reality because the advances in deep learning and the development of visual content generation methods based on models such as Variational Autoencoders (VAE) (Kingma & Welling, 2014), Generative Adversarial Networks (GAN) (Goodfellow et al., 2014) and Denoising Diffusion Probabilistic Models (DDPM) (Ho, Jain, & Abbeel, 2020). They can be used to generate dance movements or create digital choreography using existing recordings and data. This choreography can find application in education and training, where teachers demonstrate various techniques and movements to their students.

For the creation of digital Bulgarian folk choreographies, applications related to capturing the movements of the human body, creating 3D models, and building a skeleton structure (rigging) using artificial intelligence are used. They can be divided into three main categories: AI applications for motion capture (MoCap), AI applications for content creation and generation, and automated animation tools. With these applications, motion can be extracted from video, models can be created and rigged, and motion can be applied to them.

3.1. AI Applications for Motion Capture

Motion capture technology (MoCap) now makes it possible to automatically create 3D animations not only from recorded movements with special suits and cameras but also from text descriptions and videos with artificial intelligence. Motion capture from videos is primarily done by

the markerless motion capture method, which does not require placing markers at specific locations on the subject's body to track motion in 3D space. Videos are analysed using various algorithms that estimate object shape and motion, such as pose estimation algorithms (Chen, Tian, & He, 2020), which allow the extraction of pose information from the video. This data serves as the basis for trained generative models that create realistic 3D animations. This technology is essential because it has the potential to significantly reduce the time and effort required to create 3D character animation. Automating the process allows faster and more economical production of high-quality animations, which is especially useful when creating digital Bulgarian folk choreographies.

There are various methods and tools for generating 3D animation of characters from videos that use algorithms to estimate the shape and movement of objects and models that contribute to the creation of 3D animation. Among the most significant of them suitable for creating digital choreography are:

3.1.1. Rokoko

The Rokoko company was established in 2014 and is based in Copenhagen, Denmark. Its products allow users to fully capture a wide range of movements, from combat to dance scenes, without the use of body markers. The company's product is the Rokoko Vision service, which enables motion capture using AI.

This application is available in a free and paid version, works directly through the browser, and its purpose is to collect movement data from recorded video or from video captured in real time with one or two cameras. For motion data to be automatically generated, the person performing the motion must be fully in-frame with their entire body at all times. Easy-to-use Rokoko Studio editing and data integration software is available, offering cleaning filters to enhance the quality of motion capture. In addition, recordings can be exported in formats such as FBX and BVH, or live animations can be streamed directly into third-party applications such as Unreal Engine 4 and 5, Unity, and Blender via special plugins.

3.1.2. Plask

Plask is an artificial intelligence tool created by a South Korean startup. This web application analyses the movements in videos and allows their application to 3D characters. The platform has many functionalities and is designed for both beginners and professionals. It only works with single-camera footage, and performers must be positioned horizontally in front of it. Animation data can be exported in FBX and BVH formats, which are compatible with various 3D software and game engines. Machine learning technology detects and extracts performers' movements, which can be applied to standard human characters or user-defined characters. Key features listed by the company include easily capturing motion from videos without the need for expensive capture systems, through an intuitive web interface for editing and refining animations that support customisable 3D models, and collaboration features that allow teams to work together on the same project.

3.1.3. DeepMotion

In 2014, Kevin Khe founded the company DeepMotion, a pioneer in AI-based motion detection. The SayMotion application allows rapid animation generation using only a text prompt, while the Animate 3D application is designed to extract motion data from footage containing an actor performing certain actions. The resulting information can be redirected to a default 3D model provided by DeepMotion or to a model uploaded by the user. The animation can be saved by the user on his device in different formats, such as FBX, BVH, and GLB.

DeepMotion supports hand-tracking and face-scanning technologies. Multi-person tracking is one of the latest features introduced in Animate 3D that allows users to track more than one person in a single video. Depending on the subscription level, they can record the movements of up

to eight people simultaneously, including full body, hand, and face tracking, and then assign them to different 3D models.

3.1.4. Move.AI

Move.AI was founded in 2019 and is based in London, England. Move AI specialises in real-time motion capture technology, enabling users to generate high-quality 3D animations with advanced AI and computer vision without the need for costumes or specialised studios. The Move One service for iOS enables motion capture with a phone camera. Just download the app and shoot the video to get the motion data. If recording multiple people, it is necessary to use the Move Pro service.

The Move AI page contains instructions on shooting preparation, settings based on shooting volume, and calibrating the cameras. Volume Calculator shows the optimal placement of the phones that are used depending on the people being filmed. The more phones used to capture the volume, the better, especially if multiple people are going to be captured.

3.1.5. Radical Motion and Wonder Studio

Radical Motion is an AI model-based motion capture video application. RADiCAL's artificial intelligence combines deep learning, biomechanics, and computer graphics to estimate and reproduce skeleton joint rotations in 3D space using just one plain video channel. The RADiCAL Motion Live service offers instant, real-time motion capture, and the RADiCAL Motion Core service transforms pre-recorded videos into 3D motion capture. The movement of only one person at a time is tracked, so it is recommended to avoid having more than one character in the video.

Wonder Studio offers a browser-based VFX studio that not only extracts 3D animation from video shot with a single camera but also automatically adds CG characters, custom animation, and proper lighting to ensure a nice blend with the video environment.

3.2. AI Applications for Content Creation and Generation

For the creation of 3D models, graphic designers are used, who create the models by hand using traditional modelling software (Blender, Maya, and 3ds Max), use 2D/3D capture devices and reconstruction software (Meshroom and Polycam), or adjust the shape of existing humanoid templates with programs (Daz Studio, Character Creator and MetaHuman). Each of these processes requires a significant investment of time and effort. Generating 3D models with artificial intelligence greatly increases efficiency, making it easier for designers and allowing them to focus on features such as textures and animation rather than spending a lot of time modeling, capturing, or editing models.

Many tools already implement this approach to creating 3D models, which simplifies the process for both designers and users without modelling experience. One such tool is PIFuHD (Saito et al., 2020), an innovative neural network developed for 3D modelling. PIFuHD can take a single photo of a person and create a 3D model of what that person looks like.

This technology works on the principle of pixel alignment, using the information from each pixel in the image to create a detailed 3D structure. Thanks to its ability to process high-resolution images and the implicit function used, PIFuHD creates highly detailed 3D models of people.

The HyperHuman platform offers a set of tools for creating realistic 3D models of people, including animated faces, as well as for generating high-quality 3D assets. The two main tools are Rodin and Clay.

Rodin is a 3D generative model using diffusion models to automatically generate 3D digital portraits via image or text prompt, allowing text-based semantic manipulation (Wang et al., 2022). Clay is a Rodin-like generative model that supports classic text and image inputs as well as 3D controls from primitives such as multiple views, voxels, and point clouds. At its core is a large-scale generative model composed of a variational auto-encoder with different resolutions and a

minimalistic latent diffusion transformer that extracts rich 3D features directly from diverse 3D geometries (Zhang et al., 2024).

Another web platform is Meshy.ai. It uses a generative artificial intelligence model to create high-quality textures and complex 3D models from images and text queries. Its main functions include creating a texture from text, creating a texture from an image, creating a fully textured 3D model from text, and creating a textured 3D model from an image. Additionally, the platform also offers 3D model reconstruction with advanced AI features available through paid plans.

3.3. Automated Animation Tools

Automated animation tools include platforms that use algorithms for automatic rigging. They automatically generate skeleton rigs for characters, greatly reducing the time and effort required for rigging and setup. Such platforms are:

- Mixamo: Offers 3D character animation services, including an online store with downloadable animation sequences performed by various 3D characters with diverse morphologies.
- AccuRig: Specialised in generating skeleton rigs for the full body and fingers of 3D characters, exporting them in FBX and USD format for use in digital content creation (DCC) applications and game engines.

Table 1. Comparative analysis of AI video synthesis applications

Application	**Purpose**	**Degree of complexity at work**	**Need of extra hardware**	**Need of extra software**	**Price**
Rokoko	Motion capture and character animation	Medium to high (working with cameras or motion capture suits)	Specialised suits for motion capture (Rokoko Smartsuit)	Standard software (Rokoko Studio)	Subscription on a monthly basis; price on request
Plask	Motion capture using AI for animation	Medium (motion capture control interface)	Camera or smartphone with good resolution	Web-based interface, no additional software requirement	Free version, paid version with subscription
DeepMotion	3D motion capture and character animation	Medium (easy-to-use interface)	Camera (requires a good computer or smartphone with high resolution)	No additional software, uses a browser interface	Professional plan subscription; there is a free plan
Move.AI	Motion capture for professional use	High (accurate motion capture)	Camera (may require specific devices for better accuracy)	Specialised software provided by Move.AI	Price on request, meet business needs
Radical Motion	Animation and motion capture with AI	Medium (easy interface but with specific setting)	A camera or smartphone with stable video recording capability	Web interface with basic settings	Subscription on a monthly basis, there is a free plan
Wonder Studio	3D character animation with real-time AI	Medium (easy interface, automatic animation generation)	A good computer or cloud resources for data processing	Wonder Studio specialised software, browser interface	Professional plan subscription, free version
PIFuHD	3D reconstruction of human bodies from 2D images	High (requires good technical skills and setup)	High-performance computer, GPU for processing 3D data	Software for 3D reconstruction, PIFuHD demo in Google Colab	Free (exists as open source)
HyperHuman	AI to generate 3D models, assets, and animations	Low (easy interface and automatic	There are no specific hardware requirements	No need for additional software,	Free, with the option of paid

Application	Purpose	Degree of complexity at work	Need of extra hardware	Need of extra software	Price
	from text and photos	content generation)		web-based application	additional features
Meshy.ai	Automatic generation of 3D meshes from photos	Medium (image processing for 3D model)	There are no specific hardware requirements	A web-based interface for modeling and animation	Monthly subscription or indiv. packages
Mixamo	Automatic animation of 3D characters and Rigging of 3D models	Low (easy to use interface, many animation templates)	There are no specific hardware requirements	A web-based interface for animation and rigging	Free for basic features, paid add-ons
AccuRig	Rigging of 3D models and characters	Medium (automatic rigging with the possibility of settings)	There are no specific hardware requirements	Rigging program (AccuRig)	Free for basic features, paid for extra features

4. Methodology for Creating Digital Choreographies of Bulgarian Folk Dances

The creation of digital choreographies of Bulgarian folk dances requires a well-structured methodology that combines many different steps. The methodology proposed in this report combines a variety of hardware systems, software tools, and, of course, the choreographer's imagination, making it a complex but very useful and innovative approach to visualising creative solutions.

Automating the creation of a new dance would make it easier for everyone involved in it - on the one hand, choreographers would be able to materialise their ideas and see them as a finished product before presenting it to their dancers, on the other hand, performers would get a preliminary idea of their role and position in the dance.

The proposed methodology provides an opportunity for innovation and experimentation by minimising the moment of "trial-error" in the dance hall and reducing the time for construction, presentation, and assimilation of the work.

4.1. Description of the Proposed Methodology

4.1.1. Recording a combination of steps (usually corresponding to several beats of the musical accompaniment).

This step is performed by a choreographer using Motion Capture technology. The first step is recording the movements of the dancer (choreographer). This is achieved by using a suitable Motion Capture system to generate digital data about the movement of the dancer's limbs, joints, and whole body, which will then be used to drive the avatars. There are various hardware and software solutions (discussed in more detail in items 1 and 2 of this report). The choice depends on the requirements of the specific project and available resources. The generated data can often contain noises that need to be corrected in order to provide more realistic movements.

4.1.2. Creation of the 3D scene

To create the best possible quality and realistic digital choreography, it is important to create a 3D scene that matches the shape and proportions of the planned real scene as much as possible, where, in the future, real performers will reproduce the digitally created choreography. This will naturally make it easier for the dancers to quickly get an idea of their role and position in the dance and navigate the space without any problem.

4.1.3. Selection of musical accompaniment

In the fundamentals of Bulgarian folk dances lies the synchronisation of movements with the characteristic music, which is often based on complex unequal rhythms, typical of the Balkan musical culture. A variety of movement patterns are observed in the structure of the dances, ranging from simple and rhythmic steps to complex combinations of spins, jumps, and arm and body movements. When creating a choreography, the dance is divided into measures, and each element of it represents a certain number of them. It is important to keep this in mind in step one of this methodology and to record the movements based on a certain number of measures. This will facilitate synchronisation and help them to learn the individual elements of the dance more smoothly.

4.1.4. Selection of suitable 3D models (avatars)

A very important step in the methodology is the creation and customisation of the 3D models to which movements will be assigned. An additional functionality can be added - a choice of traditional folk costume corresponding to the folklore area from which the music is selected. A skeleton structure should be applied to the selected avatars. This can be achieved through a "rigging" process. Through the created skeleton, the movements of the limbs of the avatars can be controlled in order to recreate the most realistic movements.

4.1.5. Assigning the recorded step combinations and their synchronisation with one or more avatars for the specified beats

In this step, the Motion Capture imported data is applied (retargeted) to one or more avatars.

4.1.6. Synchronisation in space

Important in this step is to determine the starting position of each of the avatars in the selected 3D scene and the path direction of their movements, again for the specified number of music beats.

4.1.7. Finalisation and visualisation of the created digital choreography

In this step, it is possible to assemble all the elements of the digital choreography, make corrections in the movements of some of the 3D models, and record the result in video format.

The methodology has been tested by creating a prototype of a dancer, presented in Figure 2, using Rodin. Good results can also be achieved by PIFuHD. DeepMotion was used for motion capture, and the rigging process was realised by Mixamo.



Figure 2. A dancer prototype

5. Conclusion and Future Work

The research of technologies for capturing the movements of the human body and AI video synthesis presents new possibilities for creating digital Bulgarian folk choreographies.

The main advantages of the proposed methodology include the provision of new perspectives for the automation of the processes of creating choreographies and adapting movements, which greatly facilitates and accelerates the creative process.

Last but not least, the integration of motion capture and AI video synthesis technologies can open up new opportunities for educational and scientific initiatives related to the study of Bulgarian folk dances, as well as their presentation to future generations through innovative methods.

However, there are also challenges related to the accuracy of capturing the details of traditional movements, as well as the need to balance technological progress and the art element of the folk choreography. It is necessary to continue working on the improvement of these technologies in order to achieve the maximum result, both in terms of the quality of digital choreographies and in terms of cultural significance and respect for tradition.

References

Chen, Y., Tian, Y., & He, M. (2020). Monocular human pose estimation: A survey of deep learning-based methods. *Computer Vision and Image Understanding, 192*, 102897. https://doi.org/10.1016/j.cviu.2019.102897

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems, 27*. https://doi.org/10.48550/arXiv.1406.2661

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS 2020)*. https://doi.org/10.48550/arXiv.2006.11239

Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. *Proceedings of the 2nd International Conference on Learning Representations (ICLR 2014)*. https://doi.org/10.48550/arXiv.1312.6114

Rahul, M. (2018). Review on motion capture technology. *Global Journal of Computer Science and Technology, 18*(1). https://globaljournals.org/GJCST_Volume18/4-Review-on-Motion-Capture-Technology.pdf

Saito, S., Simon, T., Saragih, J. M., & Joo, H. (2020). PIFuHD: Multi-level pixel-aligned implicit function for high-resolution 3D human digitization. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 81–90. https://doi.org/10.1109/CVPR42600.2020.00016

Salisu, S., Ruhaiyem, N. I. R., Eisa, T. A. E., Nasser, M., Saeed, F., & Younis, H. A. (2023). Motion capture technologies for ergonomics: A systematic literature review. *Diagnostics, 13*(15), 2593. https://doi.org/10.3390/diagnostics13152593

Uzunova, Z., Chotrov, D., & Maleshkov, S. (2016). Virtual reality system for motion capture analysis and visualization for folk dance training. *Proceedings of the 12th Annual International Conference on Computer Science and Education in Computer Science*, Fulda, Germany.

Wang, T., Zhang, B., Zhang, T., Gu, S., Bao, J., Baltrušaitis, T., Shen, J., Chen, D., Wen, F., Chen, Q., et al. (2022). Rodin: A generative model for sculpting 3D digital avatars using diffusion. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Zhang, L., Wang, Z., Zhang, Q., Qiu, Q., Pang, A., Jiang, H., Yang, W., Xu, L., & Yu, J. (2024). Clay: A controllable large-scale generative model for creating high-quality 3D assets. *ACM Transactions on Graphics (TOG), 43*(4), 1–20.