Enhancing Man-Machine Interaction with Markerless Motion Capture Technology

Haotian Zhang[1]

[1] Huaer Zizhu Academy, Shanghai, 200000, China
hubery@usf.edu

Abstract. This paper explores the use of Markerless Motion Capture technology in man-machine interaction, highlighting its potential to enhance the accuracy of computer-generated outputs. The primary objective of this study is to introduce Markerless Motion Capture and demonstrate its functionality in processing images. This technology captures photos of humans from different angles and analyzes these images to form various matrices. These matrices are then applied to different equations, resulting in numerous linear functions. The study utilizes the Markerless Motion Capture dataset for analysis. Experimental results indicate that Markerless Motion Capture significantly improves the clarity and efficiency of image input in man-machine interaction. This technology has diverse applications, including medical research and game development, among others. The findings suggest that Markerless Motion Capture can play a pivotal role in advancing various fields by providing precise and efficient image processing capabilities. In the future, it is anticipated that Markerless Motion Capture will be employed in an even broader range of applications, further demonstrating its versatility and impact. The study underscores the importance of adopting advanced image processing technologies to enhance human-computer interactions and drive innovation across multiple domains.

Keywords: Markerless Motion Capture, Man-Machine Interaction, Image Processing.

1 Introduction

After ‘Industry 4.0’ a new age made of different industries, like business and politics, named ‘Industry 4.0’ provided the form of a new type of manufacturing industry start in 2011 [1]. Smart systems and services are highly relied by people who live in modern societies to support the collaboration and communication are necessary for us. Speech takes in technology, computer linguistics and vision, and artificial intelligence (AI), all of them were as results of development of software and new technologies. They promote the new application for human’s life. Man-machine interaction (MMI) in people’s life everywhere It provides a new way for human to use

Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_24

technology product Since that man-machine interaction is a key technology for both national and international research

Up to now, the information processing and visualization system can receive input and give out output to their users. This kind of interfaces help those people who is not good at using technology to use technological products more efficiently. As a result, study have introduced an "emotional dictionary", which is like a careful detective, by capturing the subtle changes in facial expressions and body language, interpreting the hidden emotional code behind it, so as to help us understand the emotional world of each user and feel their inner joys and sorrows, which can help the development of MMI applications [2]. Although the related researches have achieved some progress, the inputs of images and videos are still necessary. If account features such as speech, hand gestures or body pose can't be taken in, the stable result still cannot come out. Therefore, the main aim for this research is to talk about a new technology called Markerless Motion Capture. It is a new method to capture human behavior. It is usually used on medically. In the past years people have done a lot of things on how to capture human's movement. Movement has been a subject of rigorous scientific investigation. For instance, the Weber brothers conducted to be the earliest quantitative researches that delved into temporal and distance parameters associated with human movement. In parallel, two contemporaries, Marey and Muybridge, made significant contributions by being the first humans to employ photographic techniques in measuring human movement [3,4]. In addition, during the same period, anatomist William Brown and mathematician Otto Fisher reported data on calculating body part movements through Newtonian mechanics to measure joint forces and energy expenditure. [5]. In 1950s The California classic study helped a lot in the lives of World War II veterans. It studies the mechanics of human movement to help them regain the beauty of life [6]. Their reach provided an extensive and profound knowledge pertaining to the regular of human locomotion serves as a e fundamental basis for a technique currently employed in the research of human movement. And now, a new instrumentation and computer technologies give the chance for this technology to use in daily life.

The basic aim of this research is introducing and calculate the efficacy of Markerless Motion Capture technology in enhancing man-machine interaction through improved image processing capabilities. The research is structured to provide a comprehensive overview of how Markerless Motion Capture functions, specifically focusing on its application in generating accurate outputs from image inputs. The methodology section (Chapter 2) delves into the technical details, presenting the concepts of twists and chains, and discussing two distinct pose estimation models along with an energy model. The results of the experiments, which underscore the technology's potential to significantly improve the clarity and efficiency of image inputs, are thoroughly analyzed in Chapter 3. Ultimately, the study ends up with a key discovery point and their effects in future emphasizing the broader impact of Markerless Motion Capture in various fields such as medical research and game development.

2 Methodology

2.1 Dataset Description and Preprocessing

The operation of the Markerless motion analysis capture method depends on the current view value and does not require any preparation for the target. Main method of study is to recognize where human is and also his moving direction, and joint angle of the body through the value of view input. In large number of methods, different parts of human body are called kinetic chains that are unforced in the sense which have only one parent joint with many children [7]. Study based on previous knowledge of light and shading, this paper obtained more information about the scenes observed by the camera by fixing the joint angle limit or simulating the body motion physical [8,9]. It is as if this paper unties a string of codes, gradually reveal the mystery of the scene, let the whole picture in front of eyes vivid [10]. For example, a cyclist, his feet fixed with feet board, this creates joint confinement with feet and feet board, which is called system of closed chain or constrained motion chains.

2.2 Proposed Approach

Markerless Motion Captures can help people make more clearly, it allows the expression of geometric constraint in a label-free human pose estimation process. Compare to using traditional chain of motion constraint equations to model, Markerless Motion Captures can make model without so much limited. Traditional chain model according to the invariance of the decent upper point, however Markerless Motion Captures use the father chain with its children's Interdependencies for automatic implicit modeling and it does not need much marker-based system training [11].

Twists

The movement of the human body in a 3-dimension (3D) image can be described using certain mathematical concepts. The rotation and translation of the body are represented by matrices: the rotation matrix, and the translation matrix. Together, these matrices form a larger family known as Lie groups. For each Lie group, there is an associated vector space called the Lie algebra, which is the tangent space of the Lie group at the origin. This vector space helps researchers understand how these matrices transform in three-dimensional space. In the context of special Euclidean groups, specifically SE (3), the components of the corresponding Lie algebra can be expressed in a vector, consisting of three components. This vector is often used to describe rotational transformations and is frequently encountered in expressions describing the motion of objects. To express the current twist of a system, an exponential map is used, which helps in reconstructing the acting group. This exponential map provides a detailed understanding of how the body moves and transforms in three-dimensional space, allowing for a comprehensive analysis of the motion.

Chains

A moving chain is called as a continuous calculation of an exponential function for the twist [8]. Point on the effector at one end, the other converted from rigid body motion to the following equation:

$$Xi = exp\,(\theta\xi)\,exp\,(\theta 1\xi \cdot 1) \cdots exp\,(\theta n\xi \cdot n)\,Xi \tag{1}$$

This paper focus on a pose configuration ratio of a (6 + n) D vector to deeply explore the complex system composed of rigid body motion ξ and joint angle vectors. Specifically, the motion ξ represents the undeformed movement of the object in space, while the joint angle vector depicts the angular configuration between the joints in the system. However, in experimental setting, the exact value of vector χ, as an important part of this system, is unknown and needs to be analyzed and calculated by accurate image data. To unravel the nature of vector χ more accurately, this paper employed advanced image processing techniques to screen out key features related to vector χ from the massive image information. By analyzing these features in depth, this paper gradually narrows down the range of possible values of the vector χ, thus achieving a more accurate estimation of it. Furthermore, this study further investigated the perspective relationships in the system. This paper founds that an interdependent relationship was formed between the joint angles as the vector χ was determined. The existence of this relationship not only deepens understanding of the internal operation mechanism of the system, but also provides a strong theoretical basis for the subsequent optimization and improvement. In conclusion, this study successfully analyzed the combined relationship between the rigid body motion and the joint angle vectors, and successfully revealed the complex connection between the elements in the system, and laid a solid foundation for subsequent research [11].

Pose Estimation Model

Study performs an exhaustive extraction of the dynamic model in a specific experimental environment. This special environment not only helps us to evaluate the quality of the cone when extracting the dynamic model, but also effectively avoids the errors caused by the camera calibration as well as the foreground / background environment. During the experiment, this paper needs to find the walking test object, and for this reason, the researchers used the Poser simulator to simulate a variety of poses in a specific time period. The human 3 D view volume is between 68.07 and 67.95 liters. To construct cones of different masses, researchers used 4,8,16,32 and 64 cameras with a resolution of 640480 pixels and a horizontal viewing angle of 80 degrees. In the experimental setup, the marker-based and motion capture system consisted of eight Basler color cameras that captured images synchronously at 75 frames / s. During offline calibration, the researchers acquired parameters for both internal and external cameras and shared global reference frames. During data acquisition, images of all cameras are transmitted to multiple computers in uncompressed form. Subsequently, the computer uses a 3D point-based motion chain pose estimation algorithm to minimize the spatial distance between a given twodimensional image and the 3D model point to ensure that each image point can be reconstructed as a 3D straight line. This attitude estimation method is similar to the key-lock relationship, and each image accurately corresponds to a 3D model by accurate calculation and calibration. Just like a painting, after accurate measurement and drawing, it can be transformed into lifelike three-dimensional pictures. Furthermore, researchers employed a label-free dot motion capture system to capture whole-body motion. This unmarked:

$$(exp\,(\theta\xi)\,Xi)\pi \times ni - mi = 0 \tag{2}$$

Moreover, to minimize all correspondence, to study and optimize the function is crucial. The function (xi) represents the ordered set of joints with a direct effect on the point xi. By linearizing the exponential function by approximating exp (θξ), this paper obtains three linear equations containing 6 + n unknown variables, including six pose parameters and n joint angles. By accumulating a sufficient number of correspondences, this paper obtains a system of hyper determined linear equations to solve these unknown variables in a new squares judgment. Subsequently, the group action is reconstructed using the formula and iteratively applied to each point until convergence [11]. Moreover, to minimize all correspondence, to study and optimize the function is crucial. The function (xi) represents the ordered set of joints with a direct effect on the point xi. By linear the exponential function by approximating exp (θξ), this paper obtains three linear equations containing 6 + n unknown variables, including six pose parameters and n joint angles. By accumulating a sufficient number of correspondences, this paper obtains a system of hyper determined linear equations to solve these unknown variables in a least squares sense. Subsequently, the group action is reconstructed using the Rodrigues formula and applied to each point until convergence, as shown in the Fig. 1.

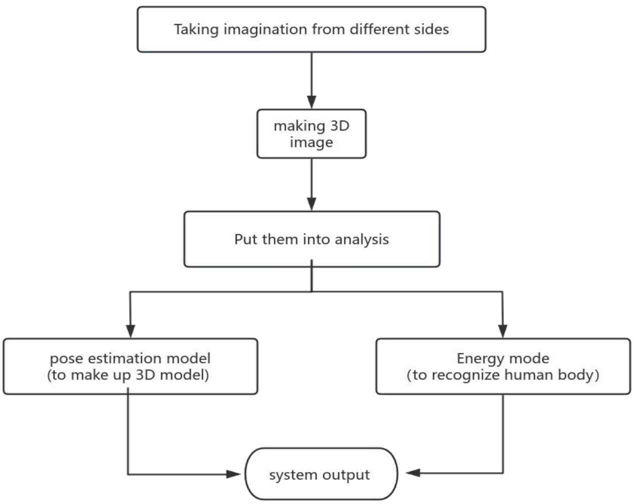

Fig. 1. Energy model overview (Photo/Picture credit: Original).

3 Results and Discussion

This paper talk about a method to make Markerless Motion Captures become useful in Man-machine interaction. This paper makes a process for how Markerless Motion Captures can make images become input for the Man-machine interaction. First, this research put large number of cameras around the tester, this paper takes photo from different sides pf tester. Then this paper makes the twists between father chain with its children's chains. Use these chains, this paper can make too different types of models. The pose estimation model can analysis the motion of human body. It takes different point on human body as an input. With this point, people can build up a real model of the human body. It also allows the assessment of the quality of the visual hull when the kinematics are extracted. After that, motion chain pose estimation algorithm based on 3D dots was used by computers. It helps computer minimize the correspondence between distance between a given plane image and the 3D model. This paper needs to ensure that each image point can be converted into a 3D straight line after reconstruction. At the same time, these three-dimensional straight lines should have their own characteristics to reflect their differences. matrix. Matrix makes it easier for computer to deal with these images. Another model is energy model. It can separate the character's image from the background. As a result, images can be Segmented into the target regions and their background. So, it helps computer recognize the environment and human more easily.

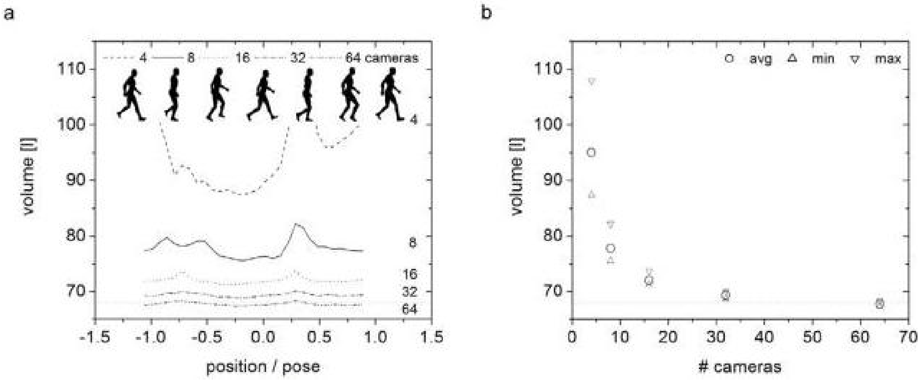

Fig. 2. The volume of human body in the image when in different state (Photo/Picture credit: Original).

As Fig. 2 shows, when human is in different type of motion and position, the volume of human body in the image change. When human is walking slowly the volume of human body is largest. While when running, the volume of human body is smallest in image. Since then, computer can analysis whether human is walking or running.

4 Conclusion

This study introduces a novel method for enhancing man-machine interaction through Markerless Motion Capture technology. This approach significantly improves the clarity and efficiency of image processing without the constraints typically associated with traditional motion capture systems. Markerless Motion Capture analyzes images and converts them into various models, aiding computers in more accurately recognizing objects within those images. Extensive experiments have been conducted to evaluate this method, and the results indicate that Markerless Motion Capture is exceptionally well-suited for man-machine interaction. By utilizing this technology, the accuracy and efficiency of computer outputs in such interactions are notably enhanced. In future research, the focus will shift towards expanding the applicability of Markerless Motion Capture to a broader range of man-machine interaction scenarios, beyond just image input. Additionally, efforts will be directed towards reducing the cost and logistical requirements of implementing Markerless Motion Capture. Currently, the technology demands a large open space and numerous cameras, which limits its usability in various settings. Therefore, it is essential to find ways to make Markerless Motion Capture more accessible and cost-effective.

References

1. Ayhun S E, Duygu T A N.: Industry 4.0 and Its Impact on Business Systems: Literature Review. Administrative, economics and social sciences, 154 (2000).
2. Tao J, Tan T.: Affective computing: A review. International Conference on Affective computing and intelligent interaction. Berlin, Heidelberg: Springer Berlin Heidelberg, 981-995 (2022).
3. Weber W, Weber E.: Ueber die Mechanik der menschlichen Gehwerkzeuge, nebst der Beschreibung eines Versuchs über das Herausfallen des Schenkelkopfs aus der Pfanne im luftverdünnten Raume. Annalen der Physik 116(1), 1-13 (1837).
4. Marey E J A M.: A Treatise on Terrestrial and Aerial Locomotion, New York: D. Appleton & Co, (1874).
5. Bell W.: Determination of the Moments of Inertia of the Human Body and its Limbs. British Journal of Sports Medicine 23(4), 258 (1989).
6. Bregler C, Malik J, Pullen K.: Twist based acquisition and tracking of animal and human kinematics. International Journal of Computer Vision 56, 179-194 (2004).
7. Sminchisescu C, Triggs B.: Estimating articulated human motion with covariance scaled sampling. The International Journal of Robotics Research 22(6), 371-391 (2003).
8. Brubaker M A, Fleet D J, Hertzmann A.: Physics-based person tracking using simplified lower-body dynamics, 2007 IEEE Conference on Computer Vision and Pattern Recognition, 1-8 (2007).

9. Balan A O, Black M J, Haussecker H.: Shining a light on human pose: On shadows, shading and the estimation of pose and shape. 2007 IEEE 11th International Conference on Computer Vision, 1-8 (2007).
10. Rosenhahn B, Brox T, Kersting U.: A system for marker-less motion capture. Künstliche Intelligenz 1, 45-51 (2006).
11. Gallier J.: Geometric methods and applications: for computer science and engineering. Springer Science & Business Media (2011).