Article

Reconstructing Three-Dimensional Human Poses: A Combined Approach of Iterative Calculation on Skeleton Model and Conformal Geometric Algebra

Xin Huang ^{1,2,*} and Lei Gao ²¹ College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300387, China² Land & Water, Commonwealth Scientific and Industrial Research Organization (CSIRO), Waite Campus, Urrbrae, SA 5064, Australia; lei.gao@csiro.au

* Correspondence: huxpiero@tjnu.edu.cn

Received: 21 December 2018; Accepted: 20 February 2019; Published: 28 February 2019

Abstract: Reconstructing three-dimensional (3D) human poses is an essential step in human body animation. The purpose of this paper is to fill the gap in virtual reality research by reconstructing postures in a high-precision human model. This paper presents a new approach for 3D human pose reconstruction based on the iterative calculation of a skeleton model and conformal geometric algebra, captured by a monocular camera. By introducing the strip information of clothes and prior data of different human limbs, the location of joint points on the human body will not be affected by the occlusion problem. We then calculate the 3D coordinates of joint points based on the proposed method of the iterative calculation of the skeleton model, which can solve the high-cost problem caused by the need for multiple cameras or a depth camera. Subsequently, we utilize high-performance conformal geometric algebra (CGA) in relation to rotation transformations in order to improve the adjustment of the postures of the human limbs. Finally, realistic 3D human poses are reconstructed—specifically, the motion of the human limbs—using a rigid transformation of CGA and a smooth connection of the limb parts based on a high-precision model. Compared with the existing methods, the proposed approach can obtain satisfactory and realistic 3D human pose estimation results using grid models.

Keywords: three-dimensional human pose estimation; skeleton model; human limb; occlusion; conformal geometric algebra; high precision human model

1. Introduction

With the continuous advance and gradual maturity of computer sciences, humans expect to obtain and deal with more information about themselves by means of computer technology, such as tracking human limb motion. As it contains personality and gait characteristics, human motion plays an important role in various fields of application, such as posture analysis and virtual reality. In the aforementioned background, three-dimensional (3D) human posture reconstruction based on videos is a popular research area [1]. 3D human posture estimation based on monocular video sequences has received more attention, owing to its advantages of low cost and less limitations. The applications have their special requirements for 3D human pose estimation, although two key performance indicators for human pose estimation algorithms are accuracy and real-time.

Recently, owing to extracted depth information, depth cameras [2,3] are applied for estimating 3D human poses and representing human activity. Kong et al. [4] presented a hybrid framework to detect joints automatically based on a depth camera. Then, 3D human poses were estimated using the located human skeleton model. Stommel et al. [5] proposed a novel method for estimating 3D human poses based on the spatiotemporal segmentation of key points, provided by depth contours, using Kinect

camera data. However, the estimation accuracy will be affected by the captured distance. Therefore, a traditional camera is used for obtaining human postures when the distance between the test people and the camera increases. Because of the absence of depth information, it is very difficult to estimate 3D human poses based on monocular video sequences. To address this challenge, a number of methods have been developed. Mehta et al. [6] proposed a real-time method to capture global 3D skeletal poses and estimate human poses based on a single RGB camera, combining a convolutional neural network with kinematic skeleton fitting. Atrevi et al. [7] extracted 3D poses using a traditional camera without any depth information, based on the correspondence between silhouettes and skeletons. Sigal et al. [8] and Babagholami–Mohamadabadi et al. [9] proposed a baseline algorithm and sparse representation to estimate 3D human poses. Furthermore, the Bayesian framework was improved by estimating a posterior distribution for sparse codes. Based on the obtained spatial and temporal features, Li et al. [10] presented an algorithm for estimating a sequence human pose in unconstrained videos. Based on the spatial model, the detection precision of body parts was improved. In order to overcome the interference to similar human poses, the corresponding depth information images were applied using several algorithms. Dinh et al. [11] presented an approach to recover 3D human poses in real-time from a depth image using principal direction analysis. Based on the introduced prior models of human poses and depth images, He et al. [12] developed a latent variable pictorial structure for estimating human poses using a monocular camera. Wu et al. [13] presented a method, called model-based recursive matching, to estimate human poses based on a depth image and 3D point cloud.

In recent years, deep learning has made considerable progress and also obtained satisfactory results in human pose estimation. Marin–Jimenez et al. [14] proposed a deep depth pose model to obtain 3D positions of body joints and reconstruct human poses. Hong et al. [15] improved traditional methods by adopting locality preserved restriction, based on a denoising auto-encoder for estimating 3D human poses. Sedai et al. [16] and Guo et al. [17] proposed a discriminative fusion method and Markov random fields, respectively, to reconstruct human poses using shape and appearance features. In order to solve the problem of occlusion, multi-view video sequences are applied for 3D human pose estimation. Sharifi et al. [18] proposed a marker-based human pose tracking and estimating method, based on particle swarm optimization, with search space partitioning.

Human pose estimation is also critical in some other areas, such as action recognition and behavior monitoring. Yang et al. [19] proposed a novel recurrent attention convolutional neural network for recognizing human action based on the sequences of video frames. Furthermore, the region of interest is visualized in order to efficiently analyze human action. Chaaraoui et al. [20] proposed a framework to recognize human behavior using multi-view cameras. Furthermore, a privacy-by-context method was used for protecting the privacy of inhabitants. Batchuluun et al. [21] recognized human behavior using camera systems, including visible light and thermal cameras. The accuracy of human behavior prediction was improved by the proposed fuzzy system.

There exist a large number of research efforts to reconstruct high-quality 3D human poses. However, the current methods suffer from the following key shortcomings: errors in human limb motion are great, the real-time reconstruction of different 3D human poses needs to be improved, and the connection between the adjacent limbs after pose adjustment is not smooth.

To address these shortcomings, we combine the iterative calculation of joint points and conformal geometric algebra (CGA) to estimate accurate 3D human poses. The images containing different human poses are captured by a camera, and the experimental 3D human models are selected from the database of free 3D models [22]. Compared with the existing work, the main contributions of this work include: (1) strip information of clothes and prior data on different human limbs are used for locating joint points, which can solve the occlusion problem between the limb part and human torso; (2) iterative calculation of the skeleton model is proposed for estimating the 3D coordinates of joint points in order to solve the high-cost problem caused by multiple cameras or a depth camera; (3) CGA makes limb motion on a 3D human model more convenient and efficient due to its obvious superiority

in rotation and transformation; and (4) a high-precision virtual human model is applied for 3D human pose estimation, which can generate more realistic and reasonable human poses.

This paper is organized as follows. In Section 2, the whole estimation process of 3D human poses is demonstrated. In Section 2.1, a 3D skeleton model and 3D limb parts are firstly introduced. The methods of locating the human joint points on a target human body and treating of the occlusion problem are then described. Subsequently, an iterative calculation of the skeleton model is applied for estimating the 3D coordinates of joint points. In Section 2.2, motion directions and the angles of various human limbs are firstly calculated by CGA. To estimate the 3D human poses, rigid transformation is then applied for adjusting the postures of limb parts on a high-precision model. In Section 3, the performance at the location on the joint points and the reconstruction performance of different human poses are analyzed. In addition, the result of 3D human pose reconstruction based on the proposed method is compared to the existing algorithms, when limb occlusion occurs. Finally, Section 4 concludes the paper.

2. The Methods

Three-dimensional human poses can be estimated based on a 3D human skeleton model and the coordinates of joint points on the human body. In the whole system, the coordinates of joint points are critical for human pose estimation, because the limb posture on the 3D human model changes according to the positions of the joint points. Therefore, the coordinates of joint points were first calculated using a color histogram and iterative calculation based on the human skeleton. Then, the limb's direction and corresponding angles of various limb parts were obtained by comparing the coordinates of joint points and the 3D skeleton model. Subsequently, a rigid transformation based on CGA was applied for adjusting the posture of the 3D human model. Finally, the realistic 3D human poses were estimated by a smooth connection on the motive limb parts. The whole process of 3D human pose estimation is shown in Figure 1.

Figure 1. The overall flow diagram of 3D human pose estimation.

2.1. Calculation of the 3D Coordinates of Joint Points on the Target Human Body

The accurate extraction of the 3D coordinates of human joint points is critical for 3D human pose reconstruction. Our method obtained the coordinates of human joint points located on the different

human motion frames by the correspondence between 2D human joint points and the 3D model's skeleton feature points.

The estimation of the 3D coordinates of human joint points was divided into three parts: first, the joint points were defined based on the biological structure of the human body and different human parts. Then, the human skeleton model can be obtained by the specific connection of the located joint points. In addition, the segmented limb parts on the human model lay a solid foundation for the limb motion. Secondly, various joint points on single-frame human motion images were located using the method of digital image processing. Finally, the focal length of the camera can be calculated by the connection model of three limb parts [23]. Therefore, the 3D coordinates of different joint points were obtained by combining the iteration method with the matching of the same points located on different human motion images.

2.1.1. Human Skeleton Model and Divided Limb Parts

The human body model can be represented by a tree stick structure [24]. As shown in Figure 2a, the whole human skeleton model is composed of various human joint points and the rigid connection parts between the adjacent joint points. All human joint points are shown in Table 1. The value g_1 is the top point of the head, and R_1 is the connection line between the head and neck parts.

Figure 2. 3D human model and its skeleton.

Table 1. The corresponding joint points of the human skeleton.

Number	g_1	g_2	g_3	g_4	g_5
Joint	Top point of head	Clavicle	Right shoulder	Left shoulder	Right elbow
Number	g_6	g_7	g_8	g_9	g_{10}
Joint	Left elbow	Right wrist	Left wrist	Waist	Right hip
Number	g_{11}	g_{12}	g_{13}	g_{14}	g_{15}
Joint	Left hip	Right knee	Left knee	Right ankle	Left ankle

Human models, applied for 3D human pose reconstruction, were selected from the free 3D model database. Various skeleton joint points on 3D human models were located by combining limb division with the appropriate proportions of the human body, and this provided an important basis for 3D human pose reconstruction based on 2D human motion images. The human model selected from the database, and the human skeleton located by the method in [25], are shown in Figure 2b,c, respectively. The major aim of the paper is to reconstruct the 3D poses of human limbs. Therefore, the human torso was considered as a whole, and the human limbs were divided into eight parts: left upper arm, left

forearm, right upper arm, right forearm, left thigh, left calf, right thigh and right calf. That is, the above divided parts were considered as a whole rigid body when the 3D human poses were estimated.

2.1.2. Joint Points' Location on the Target Human Body

The location of the joint points on human motion images is critical for 3D human pose reconstruction. In order to compare the results of 3D human pose reconstruction, we pasted the labels to locate the positions of various human joint points when human motion images were captured (see Figure 3). In the paper, the methods of manual location and automatic location were applied for locating joint points on the target human body.

Figure 3. Location on joint points on the target human body.

(1) Manual location. The accuracy of the joint points' location will have a great effect on the reconstruction results of 3D human poses. The manual location method was introduced for extracting various joint points on the target human body in order to estimate the joint points' location effect on 3D pose reconstruction. The human joint points, located using the method of manual location, and the green points correspond to various human joint points, as shown in Figure 3b. The location of various joint points can be obtained by the manual location method. That is, the method can evaluate the performance of 3D human pose reconstruction algorithms.

(2) Automatic location. As for a large number of sequence images of human motion, the method of manual location cannot be applied for all images. In addition, the automatic location of various joint points on the target human body is essential for automatic 3D human pose reconstruction. Therefore, various human joint points were extracted based on the target human motion images using the method of digital image processing. In the paper, various joint points were located by combining a color histogram with the human skeleton model.

The pasted labels were used for identifying the joint points on the target human body when human motion sequences were captured. That is, based on the method of the color histogram, the candidate pixels, with the color values of the pasted labels, are recorded to locate the human joint points. Manual location method can locate human joint points with arbitrary deviation, which can be applied for comparing the performance and robustness of 3D human pose estimation algorithm in various situations. In addition, manual location is mainly suitable for 3D human pose estimation when high-precision is needed in several applications, because some deviations will be existed when automatic location is applied.

2.1.3. Resolving the Problem of Occlusion

Three-dimensional human poses for the target human motion image can be reconstructed using a 3D human model based on the above mentioned method, when all of joint points are visible. However, some certain labels identifying human joint points were not visible due to the existing occlusion between the torso and limb or the different limb parts for some specific human poses. That is, the

occluded limb part cannot be reconstructed by the above method. Therefore, the coordinates of the occluded human joint points must be estimated in order to recover an accurate 3D pose of the occluded limb part.

In the paper, the coordinates of the occluded human joint points were estimated by combining the strip information of clothes worn by the target human body, with the measured sizes of the strip structure. The extraction of the strip information of clothes worn by the target human body is shown in Figure 4a,b. The captured human pose, when the occlusion of the limb parts occurs, is shown in Figure 4a, and the extraction result, when the algorithm of edge detection is applied, is shown in Figure 4b. As shown in Figure 4, the strip structure of the arm parts can be extracted by the edge detection method. Furthermore, the whole human arm can be divided into several parts according to the measured length sizes of each part. That is, the coordinates of the occluded joint points were estimated by the contained amount of strip parts on the visible segment of the limb. Therefore, the 3D human pose of the occluded limb can be reconstructed based on the estimated coordinates of the occluded joint points. The calculation approach is shown in Figure 4c. The points (u_0, v_0) and (u_1, v_1) are the two joint points of the occluded limb. The line segment between (u_0, v_0) and (u_s, v_s) is the visible segment, and the invisible segment corresponds to the line segment between (u_s, v_s) and (u_1, v_1) . The length sizes of line segments were measured before the experiments. R_{0s} and R_{01} are the length sizes of the visible segment and the whole limb segment. The visible proportion is defined as: $\eta = \frac{R_{0s}}{R_{01} - R_{0s}}$. Therefore, the coordinates of the occluded joint point (u_1, v_1) can be calculated by the following equation:

$$u_1 = \frac{(1 + \eta) \cdot u_s - u_0}{\eta}, v_1 = \frac{(1 + \eta) \cdot v_s - v_0}{\eta}. \quad (1)$$

(a) Human pose with occlusion.

(b) Edge detection result.

(c) The diagram of occlusion treatment.

Figure 4. The extraction of strip structures of the human arm and occlusion treatment.

Therefore, the arbitrary 3D human poses can be reconstructed by an accurate 3D pose estimation of the occluded limb parts based on the obtained coordinates of the corresponding occluded joint points.

2.1.4. Estimation of the 3D Coordinates of Human Joint Points

Data points in a 3D space were mapped onto a 2D projected plane, when the monocular camera was applied for capturing the target human poses, and this kind of transformation lead to the loss of

depth information on the data points. Therefore, the recovery of depth information on the data points was critical for 3D human pose reconstruction.

The imaging process of the camera satisfied the perspective projection model, and the principle of the model is shown in Figure 5. A and B are the two data points of the target human body in the 3D space. A' and B' are their projection points on the projected plane M . P and Q are the projection points of the data points A and B in the depth direction. According to the principle of the perspective projection, the equations are as follows:

$$s_A = \frac{x_A}{u_A} = \frac{y_A}{v_A} = \frac{OP}{f}, \quad s_B = \frac{x_B}{u_B} = \frac{y_B}{v_B} = \frac{OQ}{f}, \quad (2)$$

where x_A , y_A and OP are the X , Y and Z coordinates of point A . The values u_A and v_A are the width and height coordinates of projection point A' , respectively, and f is the focal length of the camera. The coordinates of points B and B' are similar to points A and A' . The ratios s_A and s_B are considered as scale factors of points A and B . According to the above equation, we know that when depth information changes, the scale factor will change linearly, because the parameter f remains unchanged. That is, each piece of depth information has its own scale factor. Therefore, the 3D coordinates of the data points, based on the obtained depth information, can be estimated, as long as the scale factors of the data points are calculated.

Figure 5. The perspective projection model.

To obtain the correspondence between the scale factor and the depth information of the points, the focal length f of the camera was first calculated. The calculated method of the focal length, proposed by [23], was applied in this paper in order to improve the immediacy of the whole algorithm. In the method, the focal length of the camera was calculated by combining the measured human limb length with the corresponding projection coordinates using the captured images, including the connected three limb parts.

As for 3D human pose reconstruction, the estimation of the depth information of various joint points is of the utmost importance. The limb poses of the target human body can be obtained based on the 3D coordinates of various joint points. That is, the scale factors of various joint points were calculated using the obtained focal length f . Then, the 3D coordinates of joint points can be recovered using Equation (2). The calculation process of the 3D coordinates of various joint points is as follows.

In the paper, based on the located human joint points and the skeleton model, we propose a new method, combining the iteration method with matching the same points located on different human motion images, in order to compute the scale factors of various joint points. Furthermore, different human poses are mainly reflected at the positions and directions of four limb parts, including the upper arm, forearm, thigh and calf. Therefore, it is difficult to obtain the variation of joint points of human limbs based on the captured human pose sequences, owing to the existing perspective projection. In order to overcome the disadvantages, the target human body must be located in the standard standing posture in the first frame of the motion sequential images, and all of the limb parts and torso must be parallel to the projected plane of the camera (see Figure 6). Therefore, the scale factors of the joint

points in the first image of human motion were calculated using the method of parallel projection. Based on this, the scale factors of the joint points in the other images of human motion can be obtained by matching the same joint points located on different human motion images.

Figure 6. The first frame of motion sequential images.

The lengths of human skeleton parts are critical for calculating the focal length of the camera and scale factors of joint points. Therefore, the distances between the pasted labels identifying the human joint points are measured, and the obtained data are shown in Table 2.

Table 2. The measured lengths of human skeleton parts.

Skeleton parts	R_1	R_2	R_3	R_4	R_5	R_6	R_7
Length (cm)	20	21	21	36	27	20.5	27
Skeleton parts	R_8	R_9	R_{10}	R_{11}	R_{12}	R_{13}	R_{14}
Length (cm)	20.5	24	24	43	38.5	43	38.5

In the first human motion frame, all the skeleton parts must be parallel to the camera plane. Assume that two endpoints of one skeleton part R_i are g_i and g_{i+1} . Thus, the equation is obtained as follows:

$$s_i = s_{i+1} = \frac{L_{(i,i+1)}}{L'_{(i,i+1)}} \quad (3)$$

$L_{(i,i+1)}$ is the measured length of the skeleton part R_i . $L'_{(i,i+1)}$ is the distance between the projection points, corresponding to the joint point g_i and g_{i+1} , based on the projected plane.

The 3D coordinates of the joint points in the first frame of human motion images can be obtained by the above calculation. Subsequently, the coordinates of joint points in the following images needed to be determined in order to reconstruct the poses of the human body. In the paper, the method of matching the same joint points located on different human motion images was applied in order to estimate the 3D coordinates of the joint points.

The waist point, left, and right hip points were selected as the matched feature points. In the first frame of the motion images, the coordinates of the three matched points are $\bar{g}_i(\bar{x}_i, \bar{y}_i, \bar{z}_i)$ ($i = 1, 2, 3$). The coordinates of the corresponding three matched points in the current frame of human motion images are $g_i(x_i, y_i, z_i)$ ($i = 1, 2, 3$). According to Equation (2), the following equation can be obtained:

$$x_i = \frac{u_i}{f} \cdot z_i, \quad y_i = \frac{v_i}{f} \cdot z_i. \quad (4)$$

Since the lengths of the skeleton parts remain unchanged when the human body is located in various poses, e.g., g_1 and g_2 , the objective function is defined as:

$$\begin{aligned}
f &= \left[(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2 \right] - \left[(\overline{x_{i+1}} - \overline{x_i})^2 + (\overline{y_{i+1}} - \overline{y_i})^2 + (\overline{z_{i+1}} - \overline{z_i})^2 \right] \\
&\Rightarrow f(z_i, z_{i+1}) = \left[\left(\frac{u_{i+1}}{f} \cdot z_{i+1} - \frac{u_i}{f} \cdot z_i \right)^2 + \left(\frac{v_{i+1}}{f} \cdot z_{i+1} - \frac{v_i}{f} \cdot z_i \right)^2 + (z_{i+1} - z_i)^2 \right] - S \\
&\Rightarrow f(z_i, z_{i+1}) = \frac{u_{i+1}^2 + v_{i+1}^2 + f^2}{f^2} \cdot z_{i+1}^2 + \frac{u_i^2 + v_i^2 + f^2}{f^2} \cdot z_i^2 - \frac{2(u_i u_{i+1} + v_i v_{i+1} + f^2)}{f^2} \cdot z_i z_{i+1} - S
\end{aligned} \quad (5)$$

$S = (\overline{x_{i+1}} - \overline{x_i})^2 + (\overline{y_{i+1}} - \overline{y_i})^2 + (\overline{z_{i+1}} - \overline{z_i})^2$ is a known value, which can be calculated from the first frame of the motion images. The following equations are obtained using the distances between the matched three feature points:

$$\begin{cases} f(z_1, z_2) = 0 \\ f(z_2, z_3) = 0 \\ f(z_3, z_1) = 0 \end{cases} \quad (6)$$

The above equations have three variables: z_1 , z_2 and z_3 . The Z-coordinates of the matched three feature points are selected as the initial value. Thus, a set of solutions are calculated by the iteration method, based on the Newton method. Finally, the 3D coordinates of the three matched feature points in the current frame of the motion images are obtained by the iterative solutions.

As for the different joint points in the same frame of the motion images, the scale factors of various joint points can be calculated by the iteration method, using the lengths of different skeleton parts. Assume that the scale factor of the joint point is known, and we need to calculate the scale factor of another joint point. The skeleton part connects the two joint points. Therefore, according to the principle of geometry, we can establish the following equation:

$$\begin{aligned}
&\sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 + (z_i - z_{i+1})^2} = L_{(i,i+1)} \\
&\Rightarrow (x_i - x_{i+1})^2 + (y_i - y_{i+1})^2 + (\Delta z_i)^2 = L_{(i,i+1)}^2 \\
&\Rightarrow [s_i \cdot u_i - (s_i + \Delta s_i) \cdot u_{i+1}]^2 + [s_i \cdot v_i - (s_i + \Delta s_i) \cdot v_{i+1}]^2 + (\Delta z_i)^2 = L_{(i,i+1)}^2 \\
&\Rightarrow [s_i \cdot u_i - (s_i + \frac{\Delta z_i}{f}) \cdot u_{i+1}]^2 + [s_i \cdot v_i - (s_i + \frac{\Delta z_i}{f}) \cdot v_{i+1}]^2 + (\Delta z_i)^2 = L_{(i,i+1)}^2,
\end{aligned} \quad (7)$$

where (x_i, y_i, z_i) are the 3D coordinates of point g_i , Δz_i is the difference between the depth coordinate of g_i and that of g_{i+1} , u_i and v_i are the width and height coordinates of point g_i in the projected plane, respectively. According to Equation (7), the depth difference Δz_i between point g_i and g_{i+1} can be calculated, because the parameters s_i , u_i , u_{i+1} , v_i , v_{i+1} , f and $L_{(i,i+1)}$ are known. Thus, the scale factor s_{i+1} of point g_{i+1} is as follows: $s_{i+1} = s_i + \frac{\Delta z_i}{f}$.

In summary, the 3D coordinates of various joint points in any frame of human motion images can be calculated by combining the iteration method with matching the same feature points using the lengths of skeleton parts and the coordinates of the three points in the first frame of motion images.

2.2. Limb Cooperative Motion Based on Conformal Geometric Algebra

As mentioned above, the whole 3D human models in the 3D model database [22] or those generated by some algorithms [26,27] are composed of many triangular meshes. That is, the data points and triangles on the human model will change when the movement of the limb parts on the human model occurs. It is difficult to describe the transformation process using traditional space geometry, owing to the existing complex transformation and large amount of data. Therefore, it is necessary to establish a method that can describe the above process efficiently. That is, the method must lay a solid foundation for the motion of human limb parts on a 3D human model.

CGA, which was called the generalized homogeneous coordinate [28], is a new geometric calculation representation. It can analyze and solve the problems using the language of geometry directly, because the uniform algebraic framework is already established. In addition, CGA also provides an idea for some problems associated with high technology due to its stable and efficient calculation method. Therefore, the limb cooperative motion of the human model is solved by our method using CGA, because it is greatly superior to geometric computation and treatment.

2.2.1. The Outline of Conformal Geometric Algebra

The 5D CGA is the expansion of 3D space geometry. One advantage of this algebra is that points, spheres and planes are easily represented as vectors. “Conformal” comes from the fact that it handles the conformal transformations easily. The CGA uses two additional basis vectors e_0 and e_∞ which represent the origin of coordinates and the infinity point respectively.

The three most often used products of geometric algebra are the outer, the inner and the geometric product. We used the outer product mainly for the construction and intersection of geometric objects while the inner product will be used for the computation of angles and distances. The geometric product will be used mainly for the description of transformations.

Assume that $\vec{U} = u_1e_1 + u_2e_2 + u_3e_3 + u_4e_\infty + u_5e_0$ and $\vec{V} = v_1e_1 + v_2e_2 + v_3e_3 + v_4e_\infty + v_5e_0$ are two vectors of CGA. The corresponding inner product, outer product and geometric product are defined as follows:

$$\begin{aligned} U \cdot V &= (\vec{u} + u_4e_\infty + u_5e_0) \cdot (\vec{v} + v_4e_\infty + v_5e_0) = u_1v_1 + u_2v_2 + u_3v_3 - u_5v_4 - u_4v_5 \\ U \wedge V &= (\vec{u} + u_4e_\infty + u_5e_0) \wedge (\vec{v} + v_4e_\infty + v_5e_0) \\ &= (u_1v_2 - u_2v_1)(e_1 \wedge e_2) + (u_1v_3 - u_3v_1)(e_1 \wedge e_3) + (u_2v_3 - u_3v_2)(e_2 \wedge e_3) + (u_1v_4 - u_4v_1)(e_1 \wedge e_\infty) \\ &\quad + (u_2v_4 - u_4v_2)(e_2 \wedge e_\infty) + (u_3v_4 - u_4v_3)(e_3 \wedge e_\infty) + (u_1v_5 - u_5v_1)(e_1 \wedge e_0) + (u_2v_5 - u_5v_2)(e_2 \wedge e_0) \\ &\quad + (u_3v_5 - u_5v_3)(e_3 \wedge e_0) + (u_4v_5 - u_5v_4)(e_\infty \wedge e_0) \\ UV &= U \wedge V + U \cdot V. \end{aligned} \quad (8)$$

CGA provides a great variety of basic geometric entities to compute with, namely points, spheres, planes, circles, lines and point pairs as listed in Table 3 [29,30]. They have two algebraic representations; ‘standard’ and ‘direct’. In the table, x and n indicate that they represent 3D entities by linear combinations of the 3D basis vectors e_1, e_2 and e_3 .

$$x = t_1e_1 + t_2e_2 + t_3e_3 \quad (9)$$

In the dual representation the outer product ‘ $\wedge$ ’ indicates the construction of geometric objects with the help of points x_i that lie on it. In the standard representation the meaning of the outer product is the intersection of geometric entities.

Table 3. The basic representation of conformal geometric algebra (CGA).

Geometry	Standard	Dual
Point	$P = x + \frac{1}{2}x^2e_\infty + e_0$	
Spherical surface	$S = P - \frac{1}{2}r^2e_\infty$	$S^* = P_1 \wedge P_2 \wedge P_3 \wedge P_4$
Plane	$\pi = n + de_\infty$	$\pi^* = P_1 \wedge P_2 \wedge P_3 \wedge e_\infty$
Circle	$Z = S_1 \wedge S_2$	$Z^* = P_1 \wedge P_2 \wedge P_3$
Line	$L = \pi_1 \wedge \pi_2$	$L^* = P_1 \wedge P_2 \wedge e_\infty$
Point pairs	$Pp = S_1 \wedge S_2 \wedge S_3$	$Pp^* = P_1 \wedge P_2$

2.2.2. Rotation Directions and Angles of Human Limbs

The direction vectors of various segment lines on human limbs can be obtained using 3D coordinates of their connected joint points on human motion images. Similarly, the corresponding direction vectors of various segment lines on 3D human skeleton model can also be extracted, as shown in Figure 2. Then, the rotation directions and angles on various limbs were obtained based on the human motion image and 3D skeleton model using CGA. The calculation process is described as below.

In the 5D CGA, all of the basic elements are represented as vectors (grade 1 blades), such as point, sphere, and plane. The inner product of vectors in CGA results in a scalar and can be used as a measure for distances between basic objects. In the proposed method, the distance in 3D Euclidean space can

be divided into two categories; the distance between two points, distance between point and plane, because the data points and direction sections of the human model will be used.

Assume that $\vec{T}$ is a vector of CGA. This can be written as:

$$\vec{T} = t_1\vec{e}_1 + t_2\vec{e}_2 + t_3\vec{e}_3 + t_4\vec{e}_\infty + t_5\vec{e}_0. \quad (10)$$

Therefore, the inner product of vectors $\vec{U}$ and $\vec{V}$ can be represented as follows:

$$\begin{aligned} \vec{U} \cdot \vec{V} &= (\vec{u} + u_4\vec{e}_\infty + u_5\vec{e}_0) \cdot (\vec{v} + v_4\vec{e}_\infty + v_5\vec{e}_0) \\ &= \vec{u} \cdot \vec{v} - u_4v_5 - u_5v_4. \end{aligned} \quad (11)$$

If $\vec{U}$ and $\vec{V}$ represent the two normalized points of CGA, then $\vec{U} \cdot \vec{V} = -(\vec{v} - \vec{u})^2/2$. If $\vec{U}$ and $\vec{V}$ represent the point and plane, respectively, then $\vec{U} \cdot \vec{V} = \vec{u} \cdot \vec{v} - d$.

R_i is the current skeleton line in a 3D human model, and joint points A and B are its two endpoints. R'_i is the estimated skeleton line after limb motion, which is corresponding to R_i . The position of point B remains unchanged, because it locates on rotation axis. A' is the estimated joint point A after limb motion. That is, A' and B are the two endpoints of R'_i . According to the outline of CGA, the skeleton lines R_i and R'_i can be written as:

$$L_{AB}^* = P_A \wedge P_B \wedge e_\infty, \quad L_{A'B}^* = P_{A'} \wedge P_B \wedge e_\infty. \quad (12)$$

Therefore, the rotation angle between skeleton lines R_i and R'_i can be calculated by the standardized inner product equation: $\theta = \angle(R_i, R'_i) = \arccos \frac{L_{AB}^* \cdot L_{A'B}^*}{|L_{AB}^*| \cdot |L_{A'B}^*|}$. $|L_{AB}^*|$ and $|L_{A'B}^*|$ are the length of skeleton lines R_i and R'_i .

In this part, the direction of the rotation axis was determined when the skeleton line was rotated for estimating the 3D human pose. Furthermore, the direction of the rotation axis is the normal vector of plane M , which is generated based on the two line vectors $\vec{R}_i$ and $\vec{R}'_i$.

In CGA, the plane M can be obtained based on the three points A , B and A' . The direct representation can be written as $\pi_M^* = P_A \wedge P_B \wedge P_{A'} \wedge e_\infty$. In addition, the standard representation on plane M can be expressed as $\pi_M = n + de_\infty$. The value n is the normal vector of the plane M , and d is the distance between the origin of coordinates and the plane. The above two representations can be switched using the dual operator which is indicated by $'^*$. Furthermore, the direction of the rotation axis is the normal vector n of the plane M . Therefore, the rotation direction can be obtained based on the two representations π_M^* and π_M .

2.2.3. Human Limb Motion Using Rigid Transformation

The rigid body transformation is applied for changing the posture of limb parts, because human limb parts can be regarded as a rigid body when the 3D human model is considered. A rigid body motion in 3D includes both a rotation and a translation.

In CGA, a rigid body motion of an object o is described as: $o_{Tran} = MoM^{-1}$. M is the operator of the rigid body transformation, M^{-1} is its reverse. The operator M is defined as $M = RT$. The rotor R and translator T can be expressed as follows: $R = e^{-\frac{\theta}{2}L}$, $T = e^{-\frac{1}{2}te_\infty}$. L is the rotation axis represented by a normalized vector and θ is the rotation angle around this axis, and t is the translation vector.

Assume that P is a point located on the 3D human model, and the point after the rigid body motion can be obtained as follows: $P' = MPM^{-1}$. Let x be the point corresponding to P when the human motion image is considered. Therefore, the point pairs after 3D reconstruction can be described as follows: $L_x = O \wedge x$, and O is the optical center of the camera. Furthermore, L_x and rigid body motion cannot be calculated directly, because L_x belongs to the photography space. Therefore, L_x needs to be converted by the outer product of CGA, and the line after transformation is $e_\infty \wedge L_x$.

In geometric algebra, the collinear relationships can be defined by the commutator product and the anti-exchange product. That is, we can obtain the following expression:

$X \underline{\times} L = \frac{1}{2}(XL - LX) = 0 \Rightarrow L$ and X are collinear. That is, the point P' of rigid body transformation and the point L_x of 3D reconstruction are collinear. Therefore, the following equation can be obtained:

$$P' \underline{\times} L_x = 0 \Rightarrow k((MPM^{-1}) \underline{\times} e \wedge (O \wedge x))e_{\infty} = 0, \quad (13)$$

where k is the proportional coefficient, which is used for measuring the distance of Euclidean space, and e_{∞} is applied for converting conformal space to Euclidean space.

The linearization of the transformation operator can be obtained by the operator of rigid body motion. Furthermore, the calculation process of the operator M will be simplified based on the first-order Taylor expansion formula. As for the point P , the operator M of rigid body motion can be described as:

$$\begin{aligned} MPM^{-1} &= e^{(-\frac{\theta}{2}(L' + e_{\infty}t'))} P e^{(\frac{\theta}{2}(L' + e_{\infty}t'))} \\ &\propto (1 - \frac{\theta}{2}(L' + e_{\infty}t')) P (1 + \frac{\theta}{2}(L' + e_{\infty}t')) \\ &\propto E + e_{\infty}(x - \theta(L' \cdot x) - \theta t'). \end{aligned} \quad (14)$$

If $L' = \frac{L}{\theta}$ and $t' = \frac{t}{\theta}$, then

$$MPM^{-1} = E + e_{\infty}(x - L \cdot x - t). \quad (15)$$

According to the collinear description represented by Equation (13), we can obtain the following expression:

$$L \underline{\times} (MPM^{-1}) = 0 \Rightarrow L \underline{\times} (E + e_{\infty}(x - L \cdot x - t)) = 0. \quad (16)$$

Therefore, the motion on various limb parts and 3D human pose reconstruction can be obtained based on the operator M of rigid body motion using the above linear equation.

3. Experimental Results and Validation

To test the proposed location of the joint points in the target human body images and 3D human pose reconstruction, the experiments were implemented based on the captured human body images and motion sequences for different human poses. In order to avoid the phenomenon of the labels not being able to indicate the correct joint positions when the human pose is changed, another person helped to adjust the positions of all the labels when the target human body was located in corresponding poses. That is, the human motion images were captured when all of the labels were in satisfactory positions, based on the manual operation of the second person. The 3D human model in the experiment was provided from the free 3D model database [22]. Furthermore, in order to demonstrate the location accuracy of the joint points in the target human body images, we compared the experimental results of 3D human pose reconstruction, based on the manual location, and the proposed joint point location method. In addition, the proposed 3D human pose estimation method was contrasted to the human pose reconstruction method in [23]. The whole algorithm was developed by Visual Studio 2010 and executed on an Intel CORE i5 1.7GHz PC.

All of the target human body images and human motion sequences were captured by a traditional monocular camera and mobile phone locating at a fixed position. The camera was parallel to the projected plane in the capturing process. That is, the human error in the depth information of the human joint points was eliminated based on the whole captured system. In the process of experimentation, various joint points of the target human body were identified using the pasted labels. Therefore, the lengths of various human skeleton parts and their corresponding coordinates in the human motion image can be measured based on the located joint points of the target human body. Therefore, the focal length of the camera was calculated by the connection model of three limb parts in [23] using the located clavicle point, right shoulder point, right elbow point and right wrist point.

As shown in Figure 7, subjects of the experiment needed to keep the standard standing posture, and the green points are the located human joint points. The width and height of the captured images were 1536 and 2048 pixels, respectively.

Figure 7. The location of the human joint points.

The location of the joint points of the target human body will have a great effect on 3D human pose reconstruction. In this part, the performance of 3D human pose estimation using the manual location and proposed joint point location method was tested first. In the process of experimentation, the scale factors of various joint points were calculated based on the three different joint point location groups (See Figure 8). The obtained human joint points, using the proposed automatic location method, are shown in Figure 8a. As shown in the figure, the obtained human joint points were generally located at the center of the pasted labels. It is also proved that the proposed joint point location method was effective and accurate. The first group of joint points were located by manual location, and this had little error compared to the accurately located joint points, as shown in Figure 8b. The second group of joint points, and the error between the marked joint points and the accurate joint points is very great, as shown in Figure 8c. The front view and side view of the estimated 3D human pose, using the above three groups of joint points, are shown in Figure 8e,f,h–g, respectively. As shown in the figures, the estimated 3D human pose, based on the joint points located by the proposed method, can reflect the human body pose in the motion image. However, the estimated 3D human pose using the two groups of joint points, located by manual location, had great error. In addition, the error between the estimated 3D human pose and real 3D human pose for the second group of joint points was greater than that when the first group of joint points were applied. That is, the error of the estimated 3D human pose depends on the accuracy of the location of the joint points.

Figure 8. 3D human pose reconstruction based on the different groups of joint points.

To test the accuracy of the proposed 3D human pose estimation method, different 3D human poses were estimated using the rigid motion of various limb parts based on the different human motion images captured. In this part, in order to evaluate the proposed 3D human pose estimation method, the joint points of the target human body were obtained by the method of manual location to improve the accuracy of the location of the joint points. The estimated 3D human poses using the rigid motion of various limb parts for the captured different poses are shown in Figure 9. The captured images of various human poses are shown in Figure 9a; the front view of the estimated 3D human poses is shown in Figure 9b; and the corresponding side view of the estimated 3D human poses is shown in Figure 9c. As shown in the figures, the poses in the captured images were estimated by the proposed method. Similar to Equation (17), the average errors of the 3D coordinates of the joint points for the eight human motion images were calculated. The errors on the right elbow, right wrist, left elbow, left wrist, right knee, right ankle, left knee and left ankle were 0.42, 3.12, 0.58, 4.22, 1.45, 1.52, 0.62 and 1.64 respectively. Therefore, the proposed method obtained a satisfactory 3D human pose estimation result, because all of the errors were acceptable. In addition, it also demonstrated that the rigid motion of various limb parts, based on conformal transformation, is feasible and effective.

(a) Target human body images of different poses.

(b) The front view of estimated 3D human poses.

(c) The side view of estimated 3D human poses.

Figure 9. Estimation results of different 3D human poses.

In this part, the tracking of human motion poses is studied by combining the proposed joint point location method with the 3D human pose estimation method based on the real human motion video sequences. Human motion sequence images were captured by a stationary camera, and the human error in the depth information of human joint points can be eliminated. The experimental motion sequences are the actions of the subjects in the images. The total 50 frames of a human motion

sequence were obtained under the same interval. 3D human poses were estimated based on the above sequence images using the proposed joint point location and 3D pose reconstruction methods. The 3D pose estimation results on frame 1, 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50 are shown in Figure 10. The captured images of the human motion sequence are shown in Figure 10a. The front view of the estimated 3D human poses using a 3D human model is shown in Figure 10b. The side view of the estimated 3D human poses is shown in Figure 10c in order to display the depth information of the joint points. As shown in the figures, the accuracy of the 3D pose estimation of various frames is different. Furthermore, the average errors of the 3D coordinates of the joint points for the eleven human motion images are calculated using Equation (17). The errors on right elbow, right wrist, left elbow, left wrist, right knee, right ankle, left knee and left ankle are 0.39, 3.24, 0.56, 4.34, 1.56, 1.97, 0.68 and 1.83 respectively. Therefore, as a whole, the 3D pose estimation results of human motion sequence images are satisfactory. The error in 3D human pose estimation was mainly from the inaccurate location of the joint points and the illuminant variation in the different human motion frames when the rigid motion of various limb parts is applied. In addition, compared with other skeleton parts, R_9 and R_{10} had greater errors. This phenomenon can be attributed to the accumulative errors, because the scale factors of various joint points are calculated by the iteration method.

(a) The captured human motion sequence images.

(b) The front view of the estimated 3D human poses.

(c) The side view of the estimated 3D human poses.

Figure 10. Estimation results of 3D human poses on human motion sequence images.

To estimate the accuracy of the tracking of human motion poses, the joint points obtained by the proposed automatic location method are compared to the accurate joint points based on the manual location method (see Figure 11). The width error of the located right wrist point g_7 , left wrist point g_8 , right ankle point g_{14} and left ankle point g_{15} , based on the human motion sequence images shown in Figure 10, are shown in Figure 11a. The image shown in Figure 11b corresponds to the located height error of the above joint points. As shown in the figures, the range of all the located error pixels is (3, 8).

That is, the proposed method satisfactorily locates joint points. In addition, the located errors of joint points g_8 and g_{14} have the lowest value in the 30th and 40th frames of the motion sequence images (See Figure 10). The reason is that the points of the left wrist and right ankle are nearer to the camera in the above frames. Therefore, the location of the two joint points was more accurate, because the corresponding pasted labels can be identified more easily. Furthermore, the located errors of all the joint points have the largest value, from the 5th to the 15th frame, and from the 45th to the 50th frame. In addition, as shown in Figure 11a, the right wrist point had the largest error from the 10th to the 15th frame, and from the 25th to the 30th frame. The reason is attributed to the phenomenon of the point being occluded by other human parts. Similarly, the left wrist point had the largest error from the 5th to the 15th frame. The reason is attributed to the phenomenon of the left wrist point being disturbed by the left elbow point, owing to their close positions (See Figure 10).

(a) The located width error of four located joint points. (b) The located height error of four located joint points.

Figure 11. The error in human joint points located by the proposed method.

To further estimate the location accuracy of joint points, the proposed method was compared to the five human pose estimation methods. They are those of Yang et al. [31], Chen et al. [32], Chou et al. [33], Chu et al. [34] and Luvizon et al. [35]. The PCKh, proposed by Andriluka et al. [36], was applied for measuring the location accuracy of joint points. The result is shown in Figure 12. In the lower part of the figure, the experimental 30 images, corresponding to various human poses, are presented, and in the upper part, the average location accuracy of all of the human joint points in the above 30 images is shown. As shown in the figure, apart from the specific poses, the location accuracy of the joint points was around 90%, when all of the methods are applied. Therefore, the five state-of-the-art human pose estimation methods and the proposed algorithm can obtain satisfactory location results of the human joint points. Furthermore, based on the 30 images, the average location accuracy of the joint points located by the proposed method was 93.02%. The corresponding average location accuracy was 93.23%, 93.11%, 92.78%, 93.03% and 92.04%, when the other five methods were applied. In addition, their variance in location accuracy was 9.71, 7.59, 8.47, 8.83 and 9.26, respectively. Furthermore, the variance was 7.29, when the proposed method was applied. Therefore, the proposed method had the minimal variation range of error in relation to joint point location, when various human poses were considered, although it had no superiority in location accuracy. That is, compared with the other five methods, the proposed method was the most stable algorithm. Furthermore, not all methods achieved a high-performance in relation to sitting people (e.g., poses 12, 17, 19). The reason is attributed to the phenomenon of the distances among all of the joint points decreasing when sitting people are considered, so that several joint points are often mistakenly recognized. In addition, consider the poses 13, 14, 15, 16, 17, 19, 20, 22, 23, 24, 27, 28, 29, which correspond to people with joint point occlusion. The proposed method was superior to [35] by 1.46% PCKh (91.05% vs. 89.59%), [34] by 0.44% PCKh (91.05% vs. 90.61%), [33] by 0.15% PCKh (91.05% vs. 90.90%) and [31] by 0.11% PCKh (91.05% vs. 90.94%), considering the average of the above poses. Compared with the method in [32], the proposed

method only decreased by 0.27% (91.05% vs. 91.32%). Therefore, the proposed method is feasible for joint point location when the occlusion phenomenon occurs.

Figure 12. The result of the joint point location of various human poses using different methods.

To demonstrate the variation of various limb parts in human motion sequences, the rotation angles of limb parts were extracted using the method of rigid transformation. The variation and error variation of rotation angles on the left forearm R_8 and right calf R_{12} , when the proposed method was applied based on 50 frames of motion sequence images (See Figure 10), is shown in Figure 13. The variation curve of the rotation angles, when the 3D poses of the left forearm R_8 and right calf R_{12} were estimated, is shown in Figure 13a. The rotation angle of the left forearm has the maximum value in the 10th and the 35th frame, and it has the minimum value from the 20th to the 25th frame. This indicates that the left forearm moved from the stationary state to the raised state twice. In addition, as shown in Figure 10, the two maximum values of the rotation angle correspond to the actions of right leg extension, and the minimum value corresponds to the middle action between the above two actions. The rotation angle of the left forearm in the first 15 frames was similar to that of the left forearm. In addition, the variation of the rotation angle of the right calf R_{12} remained steady in the last 30 frames, because the relative position of the right thigh and right calf remains unchanged for the action of the shot. The error variation curve of the rotation angles, when the 3D poses of the left forearm R_8 and right calf R_{12} are estimated, is shown in Figure 13b. The performance of the proposed method was satisfactory, because all of the errors were less than 8 degrees. In addition, the errors from the 30th to the 45th frame were greater than those in other frames. The reason is attributed to the phenomenon of R_8 and R_{12} being near to the screen, because the error in the location of the same 2D joint points can lead to more deviation in the 3D human pose.

Figure 13. The variation of the rotation angle using the proposed method of 3D human pose estimation.

To further prove the efficiency of the proposed 3D human pose estimation method, we compared it to the method in [23]. In this part, the set of experiments concerning the smoothness of the adjacent limb parts and the accuracy of the estimated 3D poses using the two methods are presented. The accuracy is compared based on images with the occlusion phenomenon.

The estimated results of 3D human poses based on the human motion images with occlusion, using the proposed method and the method in [23], are shown in Figure 14. As shown in the figures, the estimated 3D human motion was more realistic using the proposed method, because the algorithm of the smooth connection of the adjacent limb parts was introduced. However, the method in [23] cannot describe the realistic 3D human motion efficiently due to the distortion of the articulation when the 3D human poses were estimated. The reason is that the method in [23] only estimated the 3D poses of the limb parts and ignored the treatment of the articulated point.

In addition, as shown in Figure 14, the proposed method obtained more accurate 3D human pose estimation results. By contrast, the method in [23] cannot estimate 3D human poses with the same accuracy; especially, the deviation of the occluded limb parts was greater. This phenomenon is attributed to the estimated error in the coordinates of the occluded joint points. Due to the introduced treatment of the occluded limb parts, the proposed method solved the phenomenon of occlusion successfully and obtains a satisfactory result.

In this part, the front and side views of the estimated 3D human poses using the proposed method and the method in [37] are presented based on the image annotations in the MPII human pose dataset. As shown in Figure 15, the estimation accuracy of the 3D human parts using the proposed method was better than that using the method in [37], especially for the thigh and calf parts in model 1 and the forearms in model 2. Therefore, compared with the method in [37], the proposed method can satisfactorily estimate the human poses by calculating the 3D coordinates and CGA.

(a) The captured human motion images with occlusion.

(b) The front and side views of the estimated 3D human poses using the proposed method.

(c) The front and side views of the estimated 3D human poses using the method in [23].

Figure 14. The 3D human pose estimation with occlusion.

(a) The human motion images in the MPII Human Pose dataset.

(b) The front and side views of the estimated 3D human poses using the method in [37].

(c) The front and side views of the estimated 3D human poses using the proposed method.

Figure 15. 3D human poses estimation using the MPII human pose dataset.

The estimation accuracy can be calculated by comparing the predicted 3D coordinates of the joint points to the ground truth data. The calculated average errors of the 3D coordinates of the joint points, for the four human motion images with occlusion in Figure 14, are shown in Table 4. The error E of the joint point is defined as follows:

$$E = \sum_{i=1}^4 \frac{|x_i - x'_i| + |y_i - y'_i| + |z_i - z'_i|}{3} / 4, \quad (17)$$

where (x_i, y_i, z_i) are the coordinates of the ground truth joint points using the method of manual location in the i -th image, (x'_i, y'_i, z'_i) are the estimated coordinates of the joint points using the above two methods in the i -th image. As shown in Table 4, compared to the method in [23], there was an

important improvement in the accuracy of the location of the joint points for the proposed method, and the accuracy is critical in 3D human pose estimation.

Table 4. The average coordinate errors of the joint points using the above two methods.

Joint Points	Average Coordinate Errors of the Joint Points	
	Method in [23]	The Proposed Method
Right elbow	−2.2786	0.4721
Right wrist	−5.5375	−3.3321
Left elbow	1.5586	0.5623
Left wrist	5.8213	4.2084
Right knee	2.4268	1.4086
Right ankle	2.7365	1.5407
Left knee	2.7385	0.6074
Left ankle	4.4951	1.6587

In Table 5, the computation time of the location of the joint points, by the six methods presented in Figure 12, is shown. Compared with the other five methods, the proposed method had the lowest computation time. The computation time of the other five methods was similar. The maximum and minimum values correspond to the methods of Chen et al. [32] and Luvizon et al. [35]. Furthermore, the numbers of vertexes and computation time of various human parts, of CGA and the method in [38], are shown in Table 6. Compared with the method in [38], CGA had a longer computation time. The reason is attributed to the phenomenon that direct geometric calculation is used for changing human limb poses when the method in [38] is applied. In our method, several transformations in CGA were implemented based on the corresponding function package. That is, function references increased the whole computation time. However, computation time was acceptable and worthwhile for the whole human grid model, because the accuracy was greatly improved when the proposed method was applied. Therefore, the proposed method is not applicable for real-time human pose estimation, because grid transformation based on CGA will take some time. However, in many application areas of computer animation and human body simulation, it is not sufficient to estimate 3D human poses based solely on the human skeleton, using other methods mentioned above. The contribution of the proposed method is a high-precision human mesh model that can be used for 3D human pose estimation, so that it can be applied in many practical fields, such as virtual reality. In conclusion, compared with the other methods, the proposed method was mainly suitable for human pose estimation using high-precision 3D human mesh models and off-line processing in practical applications.

Table 5. The computation time of joint point location using the methods in Figure 12.

The Methods	Computation Time (ms)
Yang et al. [31]	60.67
Chen et al. [32]	62.36
Chou et al. [33]	61.05
Chu et al. [34]	60.47
Luvizon et al. [35]	59.45
Proposed method	47.31

Table 6. The computation time of CGA and method in [38].

Human Parts	Vertexes Numbers	Computation Time of CGA (ms)	Computation Time of Method in [38] (ms)
Right upper arm	1134	2832	216
Right forearm	719	2026	115
Left upper arm	1142	3165	206
Left forearm	721	2023	116
Right thigh	1323	3885	177
Right calf	777	2321	129
Left thigh	1322	3941	235
Left calf	776	2353	134

According to the experiments, the proposed method only estimated 3D poses for one person at present. The distance between the target person and camera was around 3 meters, and the images were captured in full light. Currently, the proposed method can predict poses with one occluded joint point. However, the method cannot be applied when the two adjacent joint points are occluded simultaneously. We will treat this phenomenon in future research.

4. Conclusions and Discussion

This paper presents an efficient algorithm for estimating 3D human poses based on iterative calculation of the skeleton model and conformal geometric algebra, using a monocular camera. The joint points of the human body are first located using the pasted labels. Strip information of clothes and prior data on the different human limbs are applied when occlusion occurs. Then, the 3D coordinates of the joint points are estimated by the method of iterative calculation, based on the obtained skeleton model. Subsequently, the motion directions and angles of various limb parts are obtained using the estimated coordinates of the joint points. Finally, the realistic 3D human poses are generated by the motion of the human limbs using rigid transformation on CGA and the smooth connection of the limb parts based on a high-precision grid model. The experimental results show that the proposed approach can obtain accurate 3D pose estimations of the human body and provide some new ideas for further research.

Future work will focus on enhancing the accuracy of joint point location and estimation using nature-inspired intelligent algorithms [39,40]. The uncertainty [41,42] of the accuracy of location and estimation will be further quantified and minimized. The future work also includes joint point location from human motion images, with no reference labels. In addition, the reality of 3D human poses in the high-precision virtual human model also needs further improvement.

Author Contributions: X.H. presented the method, collected the experimental results and drafted and revised the manuscript. L.G. conceived of the research and contributed to the analysis and revisions.

Funding: This research was funded by the National Science Foundation of China (No. 61703306), the Natural Science Foundation of Tianjin (No. 16JCQNJC00600), and the Doctoral Foundation of Tianjin Normal University (No. 52XB1302). The APC was funded by National Science Foundation of China (No. 61703306).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, X.; Wang, F.; Chen, Y. Capturing complex 3D human motions with kernelized low-rank representation from monocular RGB camera. *Sensors* **2017**, *17*, 2019. [[CrossRef](#)] [[PubMed](#)]
2. Kim, H.; Lee, S.; Lee, D.; Choi, S.; Ju, J.; Myung, H. Real-time human pose estimation and gesture recognition from depth images using superpixels and SVM classifier. *Sensors* **2015**, *15*, 12410–12427. [[CrossRef](#)] [[PubMed](#)]
3. Alazrai, R.; Momani, M.; Daoud, M.I. Fall detection for elderly from partially observed depth-map video sequences based on view-invariant human activity representation. *Appl. Sci.* **2017**, *7*, 316. [[CrossRef](#)]

4. Kong, L.; Yuan, X.; Maharjan, A.M. A hybrid framework for automatic joint detection of human poses in 110 depth frames. *Pattern Recognit.* **2018**, *77*, 216–225. [[CrossRef](#)]
5. Stommel, M.; Beetz, M.; Xu, W. Model-free detection, encoding, retrieval, and visualization of human poses from kinect data. *IEEE-ASME Trans. Mechatron.* **2015**, *20*, 865–875. [[CrossRef](#)]
6. Mehta, D.; Sridhar, S.; Sotnychenko, O.; Rhodin, H.; Shafiei, M.; Seidel, H.P.; Xu, W.; Casas, D.; Theobalt, C. VNect: Real-time 3D human pose estimation with a single RGB camera. *ACM Trans. Gr.* **2017**, *36*, 44. [[CrossRef](#)]
7. Atrevi, D.F.; Vivet, D.; Duculty, F.; Emile, B. A very simple framework for 3D human poses estimation using a single 2D image: Comparison of geometric moments descriptors. *Pattern Recognit.* **2017**, *71*, 389–401. [[CrossRef](#)]
8. Sigal, L.; Balan, A.O.; Black, M.J. HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *Int. J. Comput. Vis.* **2010**, *87*, 4–27. [[CrossRef](#)]
9. Babagholami-Mohamadabadi, B.; Jourabloo, A.; Zarghami, A.; Kasaei, S. A Bayesian framework for sparse representation-based 3D human pose estimation. *IEEE Signal Process. Lett.* **2014**, *21*, 297–300. [[CrossRef](#)]
10. Li, Q.; He, F.; Wang, T.; Zhou, L.; Xi, S. Human pose estimation by exploiting spatial and temporal constraints in body-part configurations. *IEEE Access* **2017**, *5*, 443–454. [[CrossRef](#)]
11. Dinh, D.L.; Lim, M.J.; Thang, N.D.; Lee, S.; Kim, T.S. Real-time 3D human pose recovery from a single depth image using principal direction analysis. *Appl. Intell.* **2014**, *41*, 473–486. [[CrossRef](#)]
12. He, L.; Wang, G.; Liao, Q.; Xue, J. Latent variable pictorial structure for human pose estimation on depth images. *Neurocomputing* **2016**, *203*, 52–61. [[CrossRef](#)]
13. Wu, Q.; Xu, G.; Li, M.; Chen, L.; Zhang, X.; Xie, J. Human pose estimation method based on single depth image. *IET Comput. Vis.* **2018**, *12*, 919–924. [[CrossRef](#)]
14. Marin-Jimenez, M.J.; Romero-Ramirez, F.J.; Munoz-Salinas, R.; Medina-Carnicer, R. 3D human pose estimation from depth maps using a deep combination of poses. *J. Vis. Commun. Image Represent.* **2018**, *55*, 627–639. [[CrossRef](#)]
15. Hong, C.; Chen, X.; Wang, X.; Tang, C. Hypergraph regularized autoencoder for image-based 3D human pose recovery. *Signal Process.* **2016**, *124*, 132–140. [[CrossRef](#)]
16. Sedai, S.; Bennamoun, M.; Huynh, D.Q. Discriminative fusion of shape and appearance features for human pose estimation. *Pattern Recognit.* **2013**, *46*, 3223–3237. [[CrossRef](#)]
17. Guo, C.; Ruan, S.; Liang, X.; Zhao, Q. A layered approach for robust spatial virtual human pose reconstruction using a still image. *Sensors* **2016**, *16*, 263. [[CrossRef](#)] [[PubMed](#)]
18. Sharifi, A.; Harati, A.; Vahedian, A. Marker-based human pose tracking using adaptive annealed particle swarm optimization with search space partitioning. *Image Vis. Comput.* **2017**, *62*, 28–38. [[CrossRef](#)]
19. Yang, H.; Zhang, J.; Li, S.; Lei, J.; Chen, S. Attend it again: Recurrent attention convolutional neural network for action recognition. *Appl. Sci.* **2018**, *8*, 383. [[CrossRef](#)]
20. Chaaraoui, A.A.; Padilla-Lopez, J.R.; Ferrandez-Pastor, F.J.; Nieto-Hidalgo, M.; Florez-Revuelta, F. A vision-based system for intelligent monitoring: Human behaviour analysis and privacy by context. *Sensors* **2014**, *14*, 8895–8925. [[CrossRef](#)] [[PubMed](#)]
21. Batchuluun, G.; Kim, J.H.; Hong, H.G.; Kang, J.K.; Park, K.R. Fuzzy system based human behavior recognition by combining behavior prediction and recognition. *Expert Syst. Appl.* **2017**, *81*, 108–133. [[CrossRef](#)]
22. Free 3D Models Database. Available online: http://artist-3d.com/free_3d.com/free_3d_models (accessed on 1 December 2018).
23. Zou, B.; Chen, S.; Shi, C.; Providence, U.M. Automatic reconstruction of 3D human motion pose from uncalibrated monocular video sequences based on markerless human motion tracking. *Pattern Recognit.* **2009**, *42*, 1559–1571. [[CrossRef](#)]
24. Chan, C.K.; Loh, W.P.; Rahim, A. Human motion classification using 2D stick-model matching regression coefficients. *Appl. Math. Comput.* **2016**, *283*, 70–89. [[CrossRef](#)]
25. Huang, X.; Hao, K.; Ding, Y. Human fringe skeleton extraction by an improved Hopfield neural network with direction features. *Neurocomputing* **2012**, *87*, 99–110. [[CrossRef](#)]
26. Huang, X.; Ma, X.; Zhao, Z. 3D human model generation based on skeleton segment and contours of various angles. In Proceedings of the 6th International Asia Conference on Industrial Engineering and Management Innovation, Tianjin, China, 16–18 May 2014; pp. 1033–1041.

27. Huang, X.; Zhu, Y. An entity based multi-direction cooperative deformation algorithm for generating personalized human shape. *Multimed. Tools Appl.* **2018**, *77*, 24865–24889. [\[CrossRef\]](#)
28. Zhang, Y.; Kong, X.; Wei, S.; Li, D.; Liao, Q. CGA-based approach to direct kinematics of parallel mechanisms with the 3-RS structure. *Mech. Mach. Theory* **2018**, *124*, 162–178. [\[CrossRef\]](#)
29. Zamora-Esquivel, J.; Bayro-Corrochano, E. Robot perception and handling actions using the conformal geometric algebra framework. *Adv. Appl. Clifford Algebras* **2010**, *20*, 959–990. [\[CrossRef\]](#)
30. Dorst, L.; Fontijne, D.; Mann, S. *Geometric Algebra for Computer Science: An Object-Oriented Approach to Geometry*; Elsevier: San Francisco, CA, USA, 2007.
31. Yang, W.; Li, S.; Ouyang, W.; Li, H.; Wang, X. Learning feature pyramids for human pose estimation. In Proceedings of the International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
32. Chen, Y.; Shen, C.; Wei, X.; Liu, L.; Yang, J. Adversarial PoseNet: A structure-aware convolutional network for human pose estimation. In Proceedings of the International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
33. Chou, C.-J.; Chien, J.-T.; Chen, H.-T. Self adversarial training for human pose estimation. In Proceedings of the International Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017.
34. Chu, X.; Yang, W.; Ouyang, W.; Ma, C.; Yuille, A.L.; Wang, X. Multi-context attention for human pose estimation. In Proceedings of the International Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017.
35. Luvizon, D.C.; Tabia, H.; Picard, D. Human pose regression by combining indirect part detection and contextual information. In Proceedings of the International Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017.
36. Andriluka, M.; Pishchulin, L.; Gehler, P.; Schiele, B. 2D human pose estimation: New benchmark and state of the art analysis. In Proceedings of the International Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014.
37. Chang, J.Y. DR-Net: denoising and reconstruction network for 3D human pose estimation from monocular RGN videos. *Electron. Lett.* **2018**, *54*, 70–72. [\[CrossRef\]](#)
38. Wang, C.; Ma, Q.; Zhu, D.; Chen, H.; Yang, Z. Real-time control of 3D virtual human motion using a depth-sensing camera for agricultural machinery training. *Math. Comput. Model.* **2013**, *58*, 782–789. [\[CrossRef\]](#)
39. Gao, L.; Ding, Y.; Ying, H. An adaptive social network-inspired approach to resource discovery for the complex grid systems. *Int. J. Gener. Syst.* **2006**, *35*, 347–360. [\[CrossRef\]](#)
40. Gao, L.; Hailu, A. Comprehensive learning particle swarm optimizer for constrained mixed-variable optimization problems. *Int. J. Comput. Intell. Syst.* **2010**, *3*, 832–842. [\[CrossRef\]](#)
41. Gao, L.; Bryan, B.A.; Nolan, M.; Connor, J.D.; Song, X.; Zhao, G. Robust global sensitivity analysis under deep uncertainty via scenario analysis. *Environ. Model. Softw.* **2016**, *76*, 154–166. [\[CrossRef\]](#)
42. Gao, L.; Bryan, B.A. Incorporating deep uncertainty into the elementary effects method for robust global sensitivity analysis. *Ecol. Model.* **2016**, *321*, 1–9. [\[CrossRef\]](#)