Pose Estimation of Small-articulated Animals using Multiple View Images

Sangchul Hwang
School of Mechanical Engineering
Chung-Ang University
84 Heuksukro, Dongjak-gu
156-756, Seoul, South Korea
Kinakina79@gmail.com

Young Choi
School of Mechanical Engineering
Chung-Ang University
84 Heuksukro, Dongjak-gu
156-756, Seoul, South Korea
yychoi@cau.ac.kr

ABSTRACT

The study of robots that mimic the behavior of small animals such as lizards and arthropods has been actively carried out. However, systematic analysis of the gait behavior of small animals is not prevalent in the literature. Motion analysis of most living creatures is conducted using the optical motion capture system. However, it is inapplicable to small animals owing to the difficulty of attaching optical markers. To solve this problem, markerless motion capture is being researched; however, many of these studies were performed on human subjects. Therefore, to apply a markerless motion capture system for insects that have many legs and a high degree of freedom, additional research is required. The objective of this study is to develop a system to estimate the continuous pose of small-articulated animals using a three-dimensional skeleton model of the animal from a multi-view video sequence. It includes the extraction of the extremity and root of each leg and the calculation of joint kinematics using the forward and backward reaching inverse kinematics (FABRIK) algorithm that uses the extracted extremity and root. The method developed in this study will contribute to better understand the gait behavior of small-articulated animals.

Keywords

Kinematic modeling, Markerless motion capture, Pose estimation, 3D reconstruction, 3D modeling, Inverse kinematics

1. INTRODUCTION

The research field that analyzes human and animal motion is a newly growing field of study. It is widely applied in a range of areas, from the logistics industry and space exploration to individuals’ overall living environment, such as home appliances and animation. Research on robots that resemble insects is steadily increasing. In addition, this research is expected to facilitate the collaboration of nanotechnology (NT) and information technology (IT). However, research on automatically recognizing animal motions and converting them into data, and measurement techniques by which the motions of actual animals and robots can be analytically compared, is minimal.

The objective of this study is to develop a system to track and analyze the kinematics of small animals using a three-dimensional articulated model of the animal from a multi-view video sequence. It includes the optimization of the acquisition of multi-view video images, the generation of a three-dimensional skeleton model from the multi-view images for each frame, and the calculation of joint kinematics from continuous sequences.

2. RELATED WORK

In most biomimetic robot research, the analysis of the behavior of living things is performed cognitively using two-dimensional images or videos obtained through high-speed photography or microscopy, based on biological knowledge. In addition, the difference in terms, modeling, and the research approach between biology and engineering researchers occasionally becomes an issue in the research process [Vat07].

As a solution to this problem, a markerless motion capture method was considered. The approach used in the existing markerless motion capture system is the motion reconstruction method through the interpretation of human body silhouette [Li02][Hos01].

However, it has many limitations in the reconstitution of available motion. It is inaccurate, unrealistic, and requires considerable time. Because it is affected by illumination and background changes, it is difficult to use the existing markerless motion capture system in a general environment. In order to reduce such limitations, research on markerless motion capture using a large number of cameras and a visual hull carving technique is being carried out [Mun05].

Researchers at Stanford University have developed a markerless motion capture method by obtaining the relative coordinate system for each degree of freedom from the prepared human skeleton model using human movement or motion with a series of cameras and a set of three-dimensional points from the silhouette for each individual frame [Cor06][Che03].

Gibson et al. collected the images of motions and moving behaviors of ants and spiders at various angles and traced the changes of their feature points. Research based on those traced points was conducted to form three-dimensional animations [Gib07]. However, this research was conducted focusing on animation, tracing the feature points and moving paths of the insects rather than the kinetic analysis of the minute movement of insects. Pullar et al. conducted research on a motion-tracing algorithm that uses camera images and a physics-based kinetic model to represent three-dimensional motions of spiders [Pul08].

3. EXPERIMENTAL SETUP

In order to capture the diverse motions of the small-articulated animal, six color CCD cameras (BASLER Co.'s scA640-70fc model) and a transparent acrylic box were used in this research.

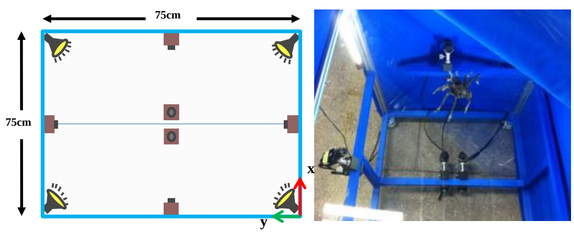


Figure 1. Captured environment that consists of a set of six calibrated cameras arranged in a blue fabric covered box

To synchronize the cameras, VIEWRUN Co.'s VR-1000 TB was used. To capture the images of the subject's motions, two below the box and four lateral cameras were utilized as shown in Figure 1. The CCD cameras can obtain an image at the speed of 70 frames per second with a 659 x 494 pixels resolution.

4. MULTI-CAMERA CALIBRATION

In general, camera calibration includes error due to the precision of patterns (chessboard), or the camera itself. To fix this problem, multiple sheets of images are obtained by diversely changing the locations of the pattern for calibration, as shown in Figure 2. Stable values are obtained by performing a filtering process at the corner points found in each of the pattern images.

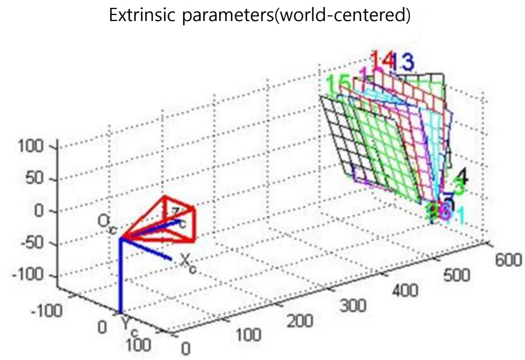


Figure 2. Camera calibration using the Camera Calibration Toolbox

In this study, the Camera Calibration Toolbox [Bou01] was used. This toolbox finds the corners from the arrangement of black and white rectangles that fit the size of the grid.

In this study, multiple cameras were used, and therefore, the accurate locations of all cameras cannot be detected through stereo calibration. The absolute position of the remaining cameras about the reference camera can be found using the relative position of a pair of cameras detected through stereo calibration.

First, multiple sheets of 2D images are obtained from the cameras, and each camera is calibrated. Next, stereo calibration is performed based on the corner points, which are back projected using the camera parameters obtained from the calibration of each camera. Based on the results, the relative position between the two cameras can be obtained. The absolute position of the remaining cameras about the reference camera is obtained by accumulating the relative position between the two calibrated cameras, as shown in Figure 3.

However, because the results of each stereo calibration contain re-projection errors, the absolute position of the camera obtained by these results includes the accumulated re-projection error. In order to solve this problem, the Lenvenberg-Marguardt (LM) minimization method is generally used. To find the slope of the errors in LM, the Jacobian matrix is established. For this research, the optimization problem was solved using sparse bundle adjustment (SBA) [Lou04][Lou09].

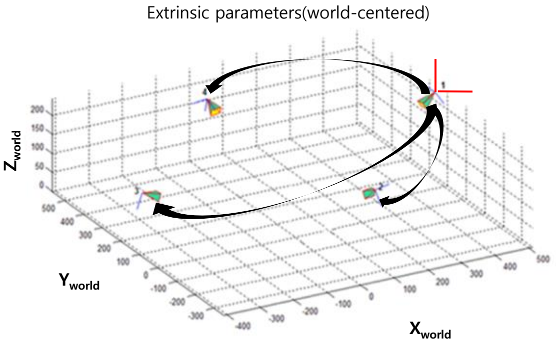


Figure 3. Absolute position of cameras obtained using the stereo calibration method

5. JOINTS POSE ESTIMATION

In this study, for articulated animals that pose difficulties in attaching markers for motion analysis, the joint position was estimated by applying the forward and backward reaching inverse kinematics (FABRIK) method [Ari11] to extremity and root(the joint between a leg and body) positions based on initial joint positions.

Initialization

The joint pose estimation of articulated animals by applying inverse kinematics was important to obtain the initial joint position. If the initial joint position is considerably different from the joint position in the second frame, an incorrect estimation is possible. Furthermore, in the case of articulated animals like tarantulas, incorrect results are more likely to occur.

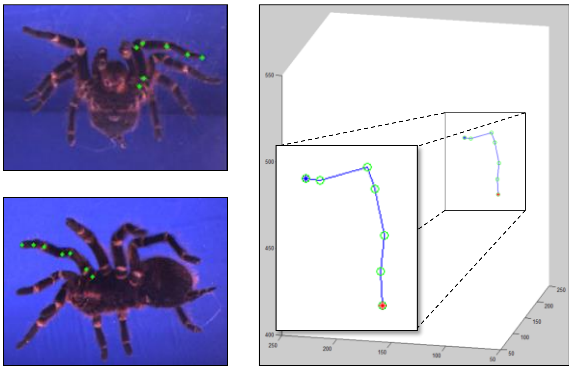


Figure 4. Three-dimensional reconstruction of the location of leg joints selected in the first frame and reconstructed in three dimensions

In addition, it is difficult to understand the length between each joint because the experimental animal is extremely difficult to control. In order to solve this problem, we manually selected the location of all joints in all images of the first frame, and reconstructed the selected locations as show in Figure 4.

In order to estimate the position of the middle joints by applying the FABRIK method, we used 16 locations of extremities and roots in the location of all joints. In addition, two reference points for root estimation were selected from the first image of continuous multiple images, as shown in Figure 5.

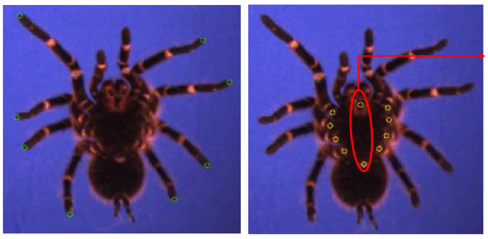


Figure 5. Selected positions of extremities, roots, and reference points for the estimation of the middle joints

Extremity estimation

In order to estimate the extremities in all frames, we used the skeletonization method, which is widely used to express the inherent structure of an object, and the skeleton pruning method based on Discrete Curve Evolution (DCE) [Bai07].

However, as shown in Figure 6(b), because of performing skeleton pruning, problems whereby the extremity is not extracted, or an unnecessary extremity is extracted, can occur.

(a) (b) (c)

Figure 6. Procedure of estimating extremities. (a) The extremities obtained from previous frame. (b) Extremities candidate extracted using the skeletonization method and the skeleton pruning method. (c) Extremity estimation result

To solve this problem, we applied a stereo image-processing method to reconstruct all the extremities obtained from multiple images in three-dimensional space. To eliminate any extremities that were reconstructed from incorrect extremity positions, we extracted those extremities that included three or more extremities in close proximity. We then grouped those extremities to obtain the centroid of each group. Next, the actual three-dimensional extremities of the experimental animal were extracted and compared with the three-dimensional extremities reconstructed from the previous frame.

Root estimation

We estimated the roots by extracting the reference point based on the assumption that root positions are always the same with respect to the reference point while moving the experimental animal. We used two images captured by the downward-facing cameras because they were well placed to detect the locations of roots.

In this study, first, an 8 × 8-pixels image is created on the next image, based on the two selected reference points in the previous image. Then, we find the coordinates close to the intensity values of the selected reference points in the previous image. Through these methods, the reference points were extracted from all frames. Next, the transformation of coordinates between lines through the two reference points in each image was calculated. Finally, the roots in all images were estimated by converting the coordinates of the selected roots in the previous image, as shown in Figure 7.

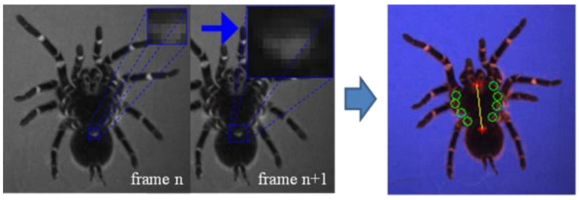


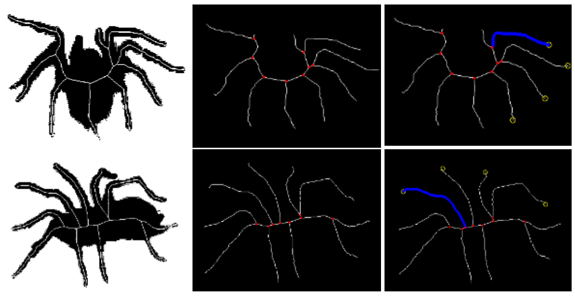
Figure 7. Procedure of estimating roots. The user selects the reference points in the frame *n* and the search algorithm for the points in the region of the frame *n+1*. The best match is indicated by the intensity value

Middle joint estimation

The position of the middle joint was estimated by applying the FABRIK method to the extremity and the root positions based on the joint position estimated from the previous frame.

However, small animals with many legs and many joints, like the tarantula in this study, generate a great deal of occulation. This makes it difficult to estimate the position of the middle joint by using general inverse kinematics methods.

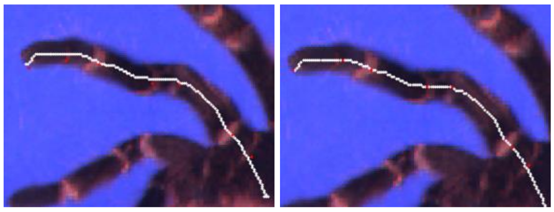
In order to solve this problem, we corrected the positions of the joints using the skeleton image generated by the silhouette image in each frame. First, we extracted only the skeletons that are applicable to the legs of the experimental animal by using the extracted points of the junction of the skeleton images, and estimated extremities in all the multiple view images, as shown in Figure 8.



(a) (b) (c)

Figure 8. Only the skeletons applicable to one leg are extracted. (a) The skeleton images. (b) The junction of the skeleton images. (c) The skeleton corresponding to one leg

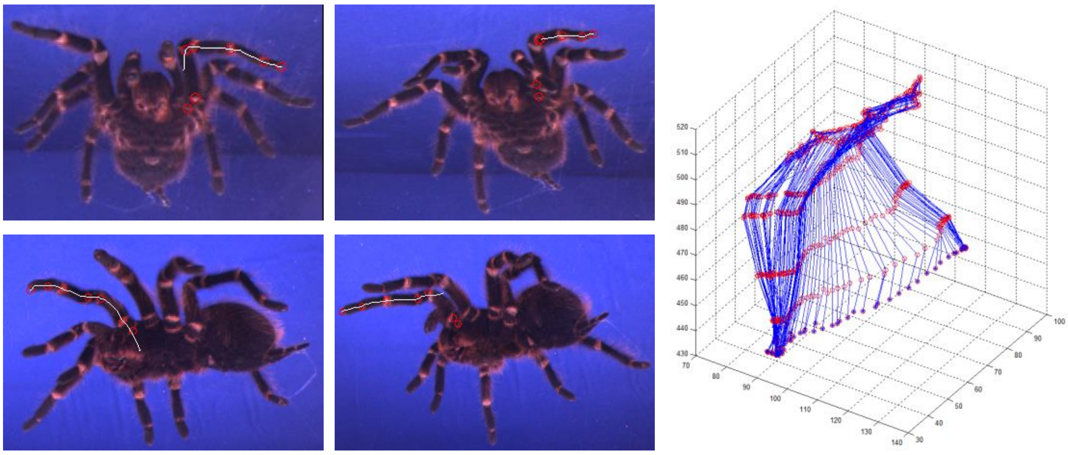
Then, after projecting the three-dimensional coordinates of the joints, the positions of which were estimated using the FABRIK method, onto the skeleton image extracted from multiple view images, the projected coordinates were corrected based on the extracted skeleton, as shown in Figure 9. We then applied stereo image processing to reconstruct the corrected coordinates. Thereafter, we grouped the three-dimensional coordinates of the joints to determine the centroid of each group. Then, by repeating this process, we extracted the optimal three-dimensional coordinates of the joints.



(a) (b)

Figure 9. Corrected position of joints. (a) The position of joints through the FABRIK method. (b) The result of correcting the position of the joints by the skeleton image

Through this method, the position of the joints can be corrected to be approximately near the skeleton of the actual target. However, this result may include an error in the process of three-dimensional reconstruction, because it is corrected using a two-dimensional image. Therefore, we made another correction according to the link lengths, and the positions of the middle joints were estimated as shown in Figure 10.



(a) (b)

Figure 10. Result of pose estimation. (a) Estimated positions of the joints in the first frame and the last frame (the hundredth frame). (b) Three-dimensional reconstruction of the joints

6. CONCLUSION

The positions of all the joints were successfully detected from multiple view images. This included the extraction of the extremities and roots of each leg, as well as the calculation of the joint kinematics using the FABRIK algorithm with the extracted extremities and roots. This is the first study to propose a means of estimating the movement of small articulated animals by using multiple view images. This has not been attempted in any previous study.

We manually selected all joint locations in all the images of the first frame and then reconstructed the selected locations. Then, we estimated the positions of the extremities by applying a skeletonization method, which is widely used to express an object's inherent structure, as well as a skeleton pruning method based on DCE. We estimated the positions of the roots based on the extracted reference points. Finally, the positions of the middle joints were estimated by applying the FABRIK method to the extremity and root positions, and then, the estimated middle joints were corrected by referring to the skeleton images. As a result, we obtained the positions of all the joints from multiple view images of the actual target. However, this result could lead to an error in a three-dimensional reconstruction because it is extracted using multiple skeleton images. This is a natural consequence of the three-dimensional reconstruction problem.

The proposed method may not be suitable for application to very small subjects, because in such cases, it is difficult to extract silhouette images. In addition, if, despite the use of multiple view images, there is excessive overlap of the joints on the images, it may be difficult to extract and then reconstruct the positions of the joints. Thus, the image processing aspect requires further study to enable the extraction of precise skeleton images.

The technique developed in this study can be adapted to implement the kinematics and optimized motion planning required for biomimetic robots.

In our future studies, we will evaluate the accuracy of the positions of the joints by comparing the positions obtained using our method with the manually selected coordinates of all joints in a series of continuous images.

7. ACKNOWLEDGMENTS

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2012R1A1A2007095) and by the Ministry of Science, ICT & Future Planning (MSIP) (No. 2007-0056094).

8. REFERENCES

[Ari11] Aristidou, A., and Lasenby, J. FABRIK: a fast, iterative solver for the inverse kinematics problem. Graphical Models, 73, No.5, pp.243–260, 2011.

[Bai07] Bai, X., Latecki, L.J., and Liu, W.Y., Skeleton pruning by contour partitioning with discrete curve evolution, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29, No.3, pp.449–462, 2007

[Bou01] Bouguet, J., Camera Calibration Toolbox for Matlab, 2004: http://www.vision.caltech.edu/bouguetj/calib_doc/ (online)

[Cor06] Corazza, S., Mündermann, L., Chaudhari, A. M., Demattio, T., Cobelli, C., and Andriacchi, T. P. A markerless motion capture system to study musculoskeletal biomechanics: visual hull and simulated annealing approach, Annals of Biomedical Engineering, 34, No.6, pp.1019-1029, 2006.

[Che03] Cheung, K. M., Visual Hull Construction, Alignment and Refinment for Human Kinematic odeling, Motion Tracking and Rendering. PhD thesis, CMU, 2003.

[Gib07] Gibson, D.P., Oziem, D.J., Dalton, C.J., and Campbell, N.W. A system for the capture and synthesis of insect motion, Graphical Models, 69, No.5, pp.231–245, 2007.

[Hos01] Hoshino, R., Yonemoto, S., Arita, D., and Taniguchi, R. Real-Time Motion Capture System Based-on Silhouette Contour Analysis and Inverse Kinematics, 7th Korea-Japan Joint Workshop on Computer Vision, 7, pp.157-163, 2001.

[Li02] Li, Yi. Hilton, A., and Illingworth, J., A Relaxation Algorithm for Real-time Multiple View 3D-tracking, Journal of Image and Vision Computing, 20, pp.841–59, 2002.

[Lou04] Lourakis, M., levmar: Levenberg-marquardt nonlinear least squares algorithms in C/C++, 2004:http://www.ics.forth.gr/~lourakis/levmar/ (online)

[Lou09] Lourakis, M. A., and Argyros, A., SBA: A Software Package for Generic Sparse Bundle Adjustment, ACM Trans. Math. Software, 36, No.1, pp. 1-30, 2009.

[Mun05] Mundermann, L., Mundermann, A., Chaudhari, A. M., and Andriacchi, T. P. Conditions That Influence the Accuracy of Anthropometric Parameter Estimation for Human Body Segments Using Shape-from Silhouette, International Society for Optics and Photonics, pp.268-277, 2005.

[Pul08] Pullar, K., and Paulin, M. Development of a model-based tracking algorithm for reconstruction of 3D spider motion, 7th joint Australia-New Zealand Mathematics Convention (ANZMC2008), 2008.

[Vat07] Vattam, S., Helms, M., and Goel, AK. Biologically-Inspired Innovation in Engineering Design: A Cognitive Study, Technical Report, Graphics, Visualization and Usability Center, Georgia Institute of Technology, 2007.