Application of Motion Correction using 3D Autoregressive Model in Kinect-based Telemedicine

Baek Seob Kim[1], *Yun Bae* Kim[2], *So-Jeong* Kim[2,*] and *Dong*-Hee Shin[3]

[1]Smart Content Research team, ETRI, 11-41 Simin-daero 327 beon-gil, Dongan-gu, Anyang, Korea,
[2]Department of Systems Management Engineering, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu University, Suwon, Korea
[3]Department of Interaction Science, Sungkyunkwan University, 25-2 Sungkyunkwan-ro, Jongro-gu, Sungkyunkwan University, Seoul, Korea

Abstract. In telemedicine, where the convergence of different types of medical treatment occurs, it is very important to establish credibility regarding the mutual communication between patients and medical workers by acquiring and sharing more accurate data. For rehabilitation treatment in particular, where motion data are required, auxiliary equipment such as a Kinect sensor is being more widely used. This study proposes a methodology for improving the motion recognition rate by compensating the noise from a Kinect sensor using a 3D autoregressive model. Moreover, this study investigates the methods applied for vitalizing the area of telemedicine under this particular trend.

1 Introduction

As people are placing increasingly high expectations on their health, the need for acquiring and recognizing relevant information is growing. Along with this growth in expectations, in the area of health communication, which helps with smoothing the interaction between patients and medical workers, the acquisition of personal health-related data and a precise delivery of the significance of such data are playing a greater role [1]. Specifically, interlinked with the development of IT technologies such as high-speed Internet, network security, video conferencing, and collaboration, telemedicine is regaining attention because it enables personally customizable health treatment, diagnosis, and remedies to be administered by connecting patients and medical workers within a network. Telemedicine is playing an appropriate role in disease protection, diagnosis, curing, and recovery by attracting the proper behavioral changes from patients based on their personal health data acquired from credible communication between the patients and medical personnel [2]. For this reason, along with proper communication between patients and medical workers, the means of acquiring and delivering data can be a crucial factor.

* Corresponding author: ai.gimso@gmail.com

With the development of information and communication technologies, the manner in which personal health data are acquired and delivered through telemedicine may be transformed. In general, medical personnel recognize, determine, and diagnose a patient's status through 2D video clips, images, and text-centric data acquired from the patient. However, patients have the responsibility to gather and deliver the data related to their own health status and conditions. However, different medical departments may require different forms of data, and the ease in acquiring such data may also vary. For instance, for the psychiatric care of anxiety or mood disorders, or the administering of internal or family medicine, which deal with chronic diseases such as high blood pressure and diabetes, telemedicine can be applied without much difficulty, and does not necessarily require support through physical contact or additionally acquired data. In contrast, rehabilitation medicine such as orthopaedics faces a difficulty when treating patients without a clear understanding of their body motions. As a way to resolve this issue, a rehabilitation program, KiReS (a kinect-based telerehabilitation system), through which a medical service provider sends a remote treatment program to a patient, allowing the patient to rehabilitate themself using a Kinect sensor, is being introduced into the market [3]. Motion recognition devices can be categorized into two types, contact-based or touch-less. Representative technologies for these two groups include OptiTrack and Kinect. OptiTrack can acquire more accurate motion data than Kinect, but is very expensive and not easily available for purchase. On the other hand, Kinect, which was originally developed for gaming but whose applications have been expanded into many areas such as medical auxiliary devices such as KiReS, has become very popular [4].

In this light, this research aims at finding a new methodology for gathering motion data in a precise manner by reducing the noise from a Kinect sensor through the application of a 3D autoregressive model. To do so, using a mathematical model, we tried to compensate the motion data in real-time by improving the motion recognition rate, and compared the experimental results with the results from a highly precise OptiTrack to see how the data acquisition can be improved. In addition, we tried to study how receiving and delivering personal health data can impact a patient's awareness and attitude toward their behavioral changes, and along with this, we tried to determine how to motivate the telemedicine industry to aggressively adopt auxiliary medical devices such as a Kinect sensor.

2 Literature Review

Health communication not only facilitates mutual communication between patients and medical personnel to improve patient health based on precise data, it also stimulates behavioral changes and active participation of the patient for the prevention, diagnosis, and treatment of diseases [1,2,5]. Therefore, the importance of health communication is receiving more attention, and its overall concept includes all processes such as information gathering, data sharing through wearable devices, and face-to-face remote counselling between medical personnel and patients. Although not regarded as significantly more effective than any other methods, this research is heavily focused on telemedicine [6], which enables the ubiquitous delivery of medical information in voice, data, and video communication formats, because owing to advancements in ICT technologies the amount of attention being paid to this field is growing.

Korea has conducted a pilot project since September of 2014 to determine the safety and validity of telemedicine. The first phase (with six-month duration) showed a high satisfaction rate of 77%. The second phase of the project started from March of 2015. Telemedicine service centers have increased in number from 18 to 50, and services have been provided for crews on deep-sea fishing vessels, soldiers, prisoners, and medically underserved people in remote areas [7]. This trend implies that telemedicine could be

highly vitalized if it can be proved that the benefits for patients and the opportunity costs for medical personnel are comparably better than those from face-to-face medical treatment.

As a friendly and heart-felt manner of interpersonal communication [8], telemedicine is said to improve our understanding of mutual interactions between patients and medical workers, and support decision-making based on behavioral changes in a more effective manner [2]. This implies that telemedicine can have an influence on the level of satisfaction regarding various medical practices. Furthermore, it is difficult to tell decisively whether telemedicine is clinically better than traditional face-to-face medical services, but its patient satisfaction has been shown to be consistently superior [9].

However, without the help of auxiliary devices in some medical departments, telemedicine experiences difficulties in achieving efficient communication, achieving medical goals, and inducing behavioral changes [2]. Without the support from auxiliary devices such as a Kinect sensor, rehabilitation therapists have had to treat patients based on their own experience and personal decision-making through physical treatments and comprehending the improvement in a patient's status with their naked eye. Clinical tests using auxiliary devices such as a Kinect sensor have proved that their use has led to a two-fold better performance compared to cases in which such devices were not used, and it was proved that a low-cost approach leads to high rehabilitation outcomes [10]. For this reason, the increasing number of applications using auxiliary medical devices such as a Kinect sensor is a natural trend and it is necessary to find a way to acquire, deliver, and share data from such devices in an efficient manner. In particular, a Kinect sensor needs to accurately recognize the motions of the patient and must be stable for use in rehabilitation treatment [10]. Medical treatment can be easily practiced when obtaining the natural body motions of the patients. Therefore, this study focuses on a way to improve the motion recognition rates by using a Kinect sensor as a telemedicine supplement for recognizing more natural body movements.

3 Methodology

3.1 Data Collection

A person has 20 Kinect-recognizable joints, two of which were selected as data acquisition targets in the present study. The two hands of the experiment participants were marked, and each participant's clapping motions were captured using both a Kinect and an OptiTrack to obtain the 3D coordinates of the target joints. A Kinect records at 30 fps, whereas an OptiTrack records at 120 fps. Thus, the amount of collected data differs between the two devices. We conducted an experiment to compensate the Kinect coordinates. Therefore, OptiTrack's frame intervals were adjusted to be the same as those of the Kinect. Specifically, data collection was conducted using a Kinect SDK 1.7 and OptiTrack Motive 1.8.1.

3.2 Data Calibration

To measure the offset in the alignment of the Kinect coordinates, it was necessary to adjust the OptiTrack, which was assumed to acquire accurate measurement data, based on the Kinect-captured 3D coordinates.

$$\widehat{x_o} = y_o(n) + mean(y_k(n)) - mean(y_o(n))$$
$$\hat{y}_o = y_o(n) + mean(y_k(n)) - mean(y_o(n)) \quad (1)$$
$$\hat{z}_o = z_o(n) + mean(z_k(n)) - mean(z_o(n))$$

Here, $x_o(n), y_o(n)$, and $z_o(n)$ are the coordinates of the n-th frame measured using OptiTrack, and $x_k(n), y_k(n)$, and $z_k(n)$ are the same coordinates as observed by the Kinect. Using formula (1), the OptiTrack coordinates were converted into Kinect coordinates.

3.3 Outlier Detection

Among the 3D coordinates observed using a Kinect, noise occurs when a person's joints are hidden behind an obstacle, or when the person is out of view of the Kinect's infrared depth vision camera; these coordinates are therefore estimated based on the coordinates of adjacent joint. However, during a small timeslot of $1/30^{th}$ of a second, some intervals may have a large amount of displacement, and after analyzing the graphs, the frames may show an obstruction of the person's natural movements. We therefore searched the outliers in all of the datasets. With the exception of those frames that contain estimated coordinates, the Inter Quartile Range (IQR) of the other frame's coordinates was calculated and the intervals outside of this range were regarded as frames with noise.

$$IQR = Q_3 - Q_1 \qquad (2)$$

Lower Bound = Q_1 – (1.5 x IQR)
Upper Bound = Q_3 + (1.5 x IQR) $\qquad (3)$

Here, Q_1 and Q_3 represent an accumulated percentile of 25% (the first quartile) and 75% (the third quartile), respectively. When the 3D coordinates observed by a Kinect sensor are smaller than the lower bound or bigger than the upper bound of equation (3), the frame belonging to the coordinates is treated as a noise-added frame.

3.4 3D autoregressive model program

When continuous observations, $\{Z_j^1, Z_j^2, Z_j^3\}_{j=1,\cdots,N} (Z_{-j}^a = Z_{N-j}^a, a = 1\cdots 3)$, are given, a 3D auto regressive model of order m can be represented through formulae (4) and (5).

$$\sum_{a=1}^{3}\frac{\sigma^a}{2}\hat{Z}_j^a = \sqrt{-1}\sum_{k=1}^{m}[\sum_{a=1}^{3}\frac{a^a}{2}Z_{j-k}^a, \sum_{b=1}^{3}\frac{\sigma^b}{2}A_k^b] = \sum_{k=1}^{m}\sum_{c=1}^{3}\frac{\sigma^c}{2}\left(Z_{j-k}\times A_k\right)^c \quad (4)$$

$$\hat{Z}_j^a = \sum_{k=1}^{m}(\mathbf{Z}_{j-k}\times A_k)^c \qquad (5)$$

The 3D coordinates observed by a Kinect have the same attributes as continuous frame-based data, namely, time-series data, and are thus suitable for an autoregressive model. We developed a program that compensates the frames containing the estimated values and the additional frames defined in Chapter 3.3 using an autoregressive model that compensates the 3D data. First, the program loads an Excel file where the 3D coordinates are stored. The frame numbers of the estimated frames that need to be compensated, and the total number (N) of frames prior to the frames to be compensated, are used as the input data. Then, order m of the model is provided to the program. After clicking the

'compensation button', vectored 3D coordinates in the autoregressive model are calculated. Finally, using formula (5), the compensated 3D values are calculated. Following these procedures, the coordinates in a frame containing noise are compensated, and the program updates the dataset to include the adjusted coordinates to compensate the following noisy frames.

3.5 Experiment Environment

To conduct the experiment on minimizing noise from a Kinect sensor, the device was placed on top of an OptiTrack to record the same motions simultaneously, and the data were compared. The lenses of the two cameras were then adjusted to look in the same direction. Given that the depth sensor camera on a Kinect is set as the origin, the 3D axes provided by Kinect SDK were set as follows. The left side of the Kinect is the positive direction of the x-axis, the right side is the negative direction of the x-axis, the upper side is the positive direction of the y-axis, the lower side is the negative direction of the y-axis, the front side is the positive direction of the z-axis, the back side is the negative direction of the z-axis, and the unit of length is measured in meters. On the other hand, the origin of the OptiTrack is set after recognizing the three markers attached to the OptiTrack origin recognition tool, and thus we aligned OptiTrack's axes with those of the Kinect before carrying out the experiments. To simplify the data observation process, the wrist of the subject was placed at the OptiTrack's origin, and we then executed the skeleton information-gathering program provided by Kinect SDK. Heuristically, we found the point where the x-axis becomes zero, and adjusted the tool's position just to the right of this point.

4 Result

We arranged the experimental environment according to the experimental methods described in chapter 3. We then collected data from the two motion recognition cameras of OptiTrack and Kinect, and converted the OptiTrack coordinates into Kinect coordinates using formula (1). Lastly, we assumed these coordinates to be the experimental values and compared them with the values from a 3D autoregressive model and based on the root mean squared error (RMSE).

Table 1. RMSE values and improvement rate after applying 3D autoregressive model

	Axis	RMSE (before)	RMSE (after)	Improvement Rate
Hand Right	X	0.023539	0.016845	28.43%
	Y	0.020090	0.018858	6.13%
	Z	0.012727	0.008986	29.39%
	Average	0.018785	0.014896	21.31%

As Table 1 shows, by comparing the experimental values with the RMSE values, the x-axis, which represents the left or right, and the z-axis, which represents the front or back, showed no distinctive differences. However, the experiments recorded an average

improvement of 21.31%. This result proves that the non-contact and touch-less camera used in a Kinect sensor can be applied to recognize natural motions when any errors are appropriately minimized.

5 Conclusion

In this study, we tried to determine whether an auxiliary device such as a Kinect sensor can be used to recognize motions more precisely when accompanied with a mathematical compensation model in the field of telemedicine, which allows patients and medical workers to communicate with each other remotely. This research proved that the motion recognition rate can be improved when the noise from a Kinect sensor is compensated using a 3D autoregressive model, and we examined certain possibilities for improving the role of health communication in the area of telemedicine.

We looked into the role and need for a Kinect sensor to be used as an auxiliary device in telemedicine based on cases taken from the industry as well as from previous studies. Being able to recognize body motions at a low cost, Kinect is a suitable device for rehabilitation treatment in terms of accuracy, processing speed, and expandability from previous researches [3-10]. Many patients seem to feel that face-to-face medical treatment is not an absolute solution for disease prevention, diagnosis, cures, and recovery, which is an attitude that is growing in magnitude. Owing to a limitation of medical laws in Korea, telemedicine is currently being carried out only for crews on deep-sea fishing vessels, soldiers, prisoners, and medically underserved people in remote areas who have very limited access to medical services; however, it is expected that when the benefits to patients such as cost reductions and improvements in medical quality increase, telemedicine will become vitalized. Moreover, the ratio of people who have a greater interest in obtaining comparative information related to their own health status will increase with the expansion of diverse health communication media, and for disease diagnosis and cures, patients will not rely solely on personnel one-to-one medical communication, but will participate equally in communication regarding their ailment with the help of additional data. In such a case, the experience and personal opinions of medical workers are important, but will not have an absolute impact on their credibility as deemed by the patients. Therefore, acquiring more objective data among medical workers and patients, along with the real-time sharing of data, is necessary, and there should be an environment that helps patients voluntarily participate in this process.

However, motions with a high level of difficulty were unable to be captured in our experiments because only one OptiTrack was used. Further researches using more than two OptiTracks to capture very difficult motions as a way to compare errors are highly recommended, and if conducted more accurate results might be made available, with a validation of the improvement rates expected. During our experiments, we realized that immersive and fun environments, along with an improvement in motion recognition, are very important factors for patients. Thus, future researches are better off applying virtual or augmented reality technologies to create a more immersive environment.

In summary, this study indicated that the proposed method for improving the motion recognition rate through a mathematical model using a Kinect sensor can be applicable to the field of telemedicine without additional costs; in addition, this method can be used to acquire and share objective data between patients and medical personnel in an effective manner. Moreover, if it is possible to elicit the participation of the patients and engage their interest by creating a virtual or augmented-reality environment that can be equipped using Kinect SDK-based 3D vision software that feels similar to real face-to-face medical

treatment, such changes will be a catalyst vitalizing the field of telemedicine.

References

1. J.M. Bernhardt, Am J Public Health, **94** (2004)
2. V.S. Freimuth, S.C. Quinn, Am J Public Health, **94,** 12 (2004)
3. D. Antón, A. Goni, A. Illarramendi, J.J. Torres-Unda, J. Seco, KiReS: A Kinect-based telerehabilitation system. *In e-Health Networking, Applications & Services (Healthcom), IEEE 15th International Conference on IEEE,* (2013)
4. B.S. Kim, *Motion Correction of 3D Autoregressive Model Using Kinect Skeletal Joint Coordinate.* Master dissertation, Sungkyunkwan University, Korea (2016). (to be published)
5. M.L. Clayman, J.A. Manganello, K. Viswanath, B.W. Hesse, N.K. Arora, J Health Commun, **15,** (2010)
6. D.J. Kim, Korea Institute for Health and Social Affairs Issue & Focus, **270** (2015)
7. C.S. Park, *Report* (*Korea Ministry of Health and Welfare*, 2015)
8. A. Duggan, Journal *of health communication,* **11,** (2016)
9. W.D. Leach, *If You Bill It, They Will Come: A literature review on clinical outcomes, cost-effectiveness, and reimbursement for telemedicine,* (California Telemedicine and eHealth Center, Sacramento, CA, 2009)
10. H. Mousavi Hondori, M. Khademi, Journal of Medical Engineering, (2014)