1. INTRODUCTION

Knee surgery is operated to recover the normal functionality of the knee. Knee replacement surgery is an operation to replace the damaged knee due to osteoarthritis (OA) or rheumatoid arthritis (RA) [1] with artificial one. The number of cases in Japan exceeds 80,000. It has increased year by year, and becomes one of the major orthopaedic surgeries [2].

Knee replacement surgery is mainly classified into total knee replacement arthroplasty (TKA) and unicompartmental knee arthroplasty (UKA). TKA replaces all knee joint surfaces, while UKA replaces a part of the knee joint surface with mild damage. The surgical procedure is quite complicated (it has roughly 27 phases), and it uses many surgical instruments (over 120 kinds), and hand assembling of the instruments is also required during the operation. Most hospitals use multiple models, and the surgical procedure and the surgical instruments are different among the models.

For instrumental nurses who are in charge of multiple kinds of surgical operations, it is a heavy burden to recognize these complicated procedures and surgical equipment [3]. It might cause a serious incident leading to surgical errors, operation time extension, poor prognosis, deterioration in the quality of surgery due to unfamiliar instrumentation, and so on. Therefore, it is desired to develop a real-time intraoperative navigation system that supports instruments nurses.

We introduce a computer-aided orthopaedic surgery (CAOS)-AI navigation system for operating room nurse. One of the principal function of CAOS-AI navigation system is to recognize the current phase of orthopaedic procedures from surgeon-wearable video camera images. The method plays the fundamental role of CAOS-AI navigation system. The proposed method is based on a convolutional-long short-term memory (LSTM) network. We also investigate the efficient CNN model in some competitive models such as VGG16, DenseNet, and ResNet to improve the recognition accuracy. Experimental results in unicompartmental knee arthroplasty (UKA) surgeries showed that the proposed method achieved a phase recognition accuracy with 48.2%, 41.2%, and 53.6% using VGG16, DenseNet, and ResNet, respectively.

Keywords: Computer-aided orthopaedic surgery, CAOS-AI navigation system, Operating room nurse, Surgery video, Surgical phase recognition, Deep learning, Wearable camera
have been used. Thus, the existing methods cannot be applied to the knee arthroplasty surgery directly.

In Kansei engineering, image processing using AI is one of the important technologies. Advanced image processing has a potential to be replaced the human sense. The human sense has the ability to go beyond existing engineering techniques. Kansei engineering is a technique that introduces human sense into technology. Surgical navigation system based on human sense are expected to have higher capabilities than before.

In this paper, we propose surgical navigation system using wearable video camera (smart glasses) based on human sense and surgical recognition method required that system. We introduce a convolutional-long short-term memory (LSTM) network into surgical phase recognition. The proposed method recognizes phases from knee arthroplasty surgical video images taken by using surgeon wearable video camera.

This paper consists of the following sections. Section 2 describes materials used in this study. Section 3 proposes a model to recognize the current surgical phase. Section 4 shows experiment results and discusses the outcome. Finally, the study is concluded with future scope in Section 5.

2. PRELIMINARIES

2.1 Unicompartmental Knee Arthroplasty

UKA is a surgical operation which replaces a part of the damaged knee joints with artificial one. This study employed a UKA produced by Zimmer Biomet G. K. (US). The procedure consists of 27 phases [8] as tabulated in Table 1. Among them, this paper focuses on 11 phases, Phase 1, 3, 6, 7, 8, 9, 12, 13, 16, 18, and 27 in order to validate the fundamental performance of the proposed method because they appeared frequently in the surgical video.

Phase 1 and Phase 2 make incision into the knee skin using a scalpel and a scissor, respectively. Phase 3 makes incision into the interior skin using ultra solidifying and incision apparatus. Phase 6 inserts a tibial cutting guide. Phase 7 and Phase 8 perform the horizontal and the vertical resection of the tibia using a blade, respectively. Phase 10 inserts a drill guide to support femoral drilling, and Phase 12 drills the femoral bone. Phase 13 resects the femur. Phase 16 measures the gap between the modified femoral and the tibial bones. Phase 18 cuts the femoral condyle using a spherical cutter, and finally Phase 27 sutures.

| Phase | Procedure |
|-------|-----------|
| 1     | First incision using a scalpel |
| 2     | Second incision using scissors |
| 3     | Third incision using ultra solidifying and incision apparatus |
| 4     | Osteophyte excision |
| 5     | Insert a sizing spoon |
| 6     | Insert a tibial cutting guide |
| 7     | Tibial plateau horizontal resection |
| 8     | Tibial plateau vertical Resection |
| 9     | Insert a rod |
| 10    | Insert a drill guide |
| 11    | Link the rod with the drill guide |
| 12    | Femoral drilling |
| 13    | Femoral resection |
| 14    | Meniscal resection |
| 15    | Milling |
| 16    | Measure the gap |
| 17    | Insert a preventing impingement |
| 18    | Sverhriel Cutter |
| 19    | Chisel the back area |
| 20    | Final preparation of the tibial plateau |
| 21    | Impact the tibial component using a tibial impactor |
| 22    | Final trial reduction |
| 23    | Drill a cement hole |
| 24    | Insert paste cement to the cement hole |
| 25    | Insert a tibial component |
| 26    | Insert a femoral component |
| 27    | Sutures |

2.2 Knee Arthroplasty Surgeon Wearable Video Camera Data

This study employed a smart glass for surgical video streaming system proposed in Hiranaka et al. [9]. The smart glasses is a small wearable electronic device with an optical head-mounted display. The principal function includes video recording, wireless connectivity with bidirectional communication, hands-free operation, augmented reality (AR), and so on. And, it is lighter than usual head mounted camera. While the smart glasses has been used by many industrial applications, there are very few cases used in CAOS in which the smart glasses is worn by a surgeon.

The whole process of UKA surgery was recorded by using a smart glasses (InfoLinker, West Unititis Co. Ltd., JAPAN), and was archived in a private database. This study was approved by the local Ethics committee of Takatsuki Hospital (Takatsuki, Japan).

This study acquired six UKA videos. Table 2 shows the description of orthopaedic surgical videos. The dataset consists of heteronomous videos including different frame rates and sizes. Example of frames in the UKA surgery is shown in Figure 1. Poor room illumination, tiny and similar surgical instruments and constrained environment make the orthopedic surgical recognition task difficult in
In preliminary, 218 video clips taking the surgical phases of interest are cropped, and only the cropped video clips are analyzed in this paper.

### 2.3 Long Short Term Memory Network (LSTM)

To recognize the surgical phase from video data, SV-RC Net [10] and B. Namazi et al. [11] have used LSTM [12] with CNN. The LSTM is one of the recurrent neural networks, and has been applied to analyze time-series data such as stock prices, speech recognition, text generation from image, language, language translation, etc.

The architecture of single LSTM is illustrated in Figure 2. The LSTM unit has three gates: input gate $i_t$, forget gate $f_t$, and output gate $o_t$, to modulate the interactions between the memory cell $c_t$. The input gate controls the degree of effect by the memory cell to the current hidden state. The details of LSTM follow Ref. [12].

### 3. PROPOSED METHOD

#### 3.1 Computer-aided Orthopaedic Surgery-AI Navigation System

The overview of proposed CAOS-AI navigation system is shown in Figure 3. As denoted by a red square, it first estimates the current surgical phase from video camera images acquired by surgeon’s smart glasses. This paper proposes a method to perform the real-time phase recognition. Using the estimated surgical phase, the system instructs an operation nurse to prepare surgical tools used in the next phase via smart glasses in real time. Preliminary results were shown in our previous work [7].
3.2 Surgical Phase Recognition Based on Convolutional-LSTM Network Model

Figure 4 shows the proposed phase recognition model. As shown in Figure 4, the LSTM locates between a CNN (to extract image feature) and a classification layer including fully connected layer and output layer (Softmax). The network structure is summarized in Table 3.

At time Phase $x_t$, $N (= 10$ is used in this study) previous frames, $x_{t-10}$, $x_{t-9}$, ..., $x_t$, with 224x224 pixels are feed to a CNN described in the next section. Although the proposed convolutional LSTM model includes $N (=10)$ CNNs, the same trained CNN is used. The CNN extracts 512-D feature vector from each frame, and the extracted feature vector is given to LSTM network. Finally, Softmax layer outputs a vector with probability of 11 classes (i.e., phases). The class with the highest probability is selected as the most probable surgical phase.

CNN model and LSTM model are trained simultaneously based on image information and temporal information of images. One video clip consists of 10 frames, and the phase number is labeled to only 10th frame. Therefore, our model outputs the class provability vector with respect to the previous $N$-1 frames. The obtained vector is used to calculate the loss function. All networks including CNN and LSTM are trained using the calculated result of loss function at the same time.

3.3 Convolutional Neural Network for Surgical Phase Feature Extraction

The proposed method uses CNN to extract features from each frame. The CNNs evaluated are VGG16 [13], MobileNet [14], Dense Net [15], and ResNet-50 [16]. After training the models using the Image Net [17], the output layer including fully-connected and Softmax is replaced with connected layer, LSTM network, new fully-connected layer and new Softmax layer. As an example, the structure of VGG16 is shown in Figure 5 and Table 4.

A single frame given to the input layer has 224x224 pixels with RGB color channel. Then, 512-D feature vector is obtained from the connected layer of CNN.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Training Details and Evaluation Metrics

The proposed method was implemented by using Python with Keras deep learning library as Tensorflow backend using a NVIDIA GTX 1080 accelerator. The hyper parameters in the network were; LSTM dropout rate = 0.5, the learning rate for Adam [18] optimizer of our model was 0.00001. Also, we used categorical cross entropy as loss function. In this study, we used epochs of 50 in all training process in order to compare the performance of VGG16, DenseNet, ResNet, and MobileNet.

To evaluate the trained model, we used accuracy formulated by Eq. (1). Total # of video clips is the number of all frames used for testing. # of correctly recognized video clips is the number of frames whose estimated phase is equal to the true label (surgical phase label).
4.2 Experimental Results

In our experiment, we used 6 surgical videos. We evaluated the proposed method according to leave-one-out-cross-validation (LOOCV) where five video data were used for training and the remained one was used for evaluation. Each video included 11 surgical phases (i.e., classes). In preliminary, we have manually extracted 218 surgical video clips. One video clip consists of 10 frames. The data balance among classes is important to improve the performance. The number of video clips at each phase is shown in Figure 6. The average number of video clips was 19.9.

Table 5 summarises the experimental results. The results of convolutional-LSTM using MobileNet and ResNet-50 were 53.6% and 53.7%, respectively, and they were better than the others. Also, the average accuracy of video 3 was the highest in all experiments, and the average accuracy of all models was 49.2%.

The learning curve of the accuracy and the loss during the training process is shown in Figure 7. Figure 7(b) indicates that MobileNet and ResNet-50 were lower than the others, and DenseNet and VGG16 were overfitting because the loss increased over 20 epochs. Also, the loss of MobileNet did not converge yet at 50 epochs. We have to increase the epoch to obtain the proper performance of MobileNet.

Table 6 shows the confusion matrix of phase recognition result using ResNet-50. The accuracy of Phase 27 was the highest because the surgeon’s hand motion was quite different from the other phases. And, the accuracy of phase 3, 9 were low because the number of video clip were smaller than the others. The imbalanced dataset could affect misclassification.

![Figure 6: Number of video clips per phase](image)

\[
\text{Accuracy} = \frac{\text{# of correctly recognized video clips}}{\text{Total # of video clips}} \quad (1)
\]

![Table 5: Testing accuracy on TKA and UKA dataset with LOOCV](table)

![Figure 7: Learning curve during training](image)
5. CONCLUSION

This paper has proposed a CAOS-AI navigation system based on convolutional-LSTM. It included CNN models such as VGG16, ResNet-50, DenseNet, and MobileNet. The experimental results with LOOCV showed that we achieved an average accuracy of 53.7% by using MobileNet even though the training dataset was small. Also, we found that it was effective to adopt convolutional-LSTM to orthopedic surgery video recorded by wearable camera (smart glasses).

In future, we need to increase the number of data to improve the performance of our system. We are also going to apply data augmentation techniques, and to find optimal parameters in the model. And, we will extend the proposed method to evaluate the order of surgical phases using HHMM [19] and PKI function [9].

ACKNOWLEDGEMENTS

This work was supported in part by KAKENHI JSPS Grant-in-Aid grant (18F18377).

REFERENCES

1. Murray, D. W., et al.; A randomised controlled trial of the clinical effectiveness and cost-effectiveness of different knee prostheses: the Knee Arthroplasty Trial (KAT), Health Technology Assessment, 18(19), pp.1-235, 2014.

2. Yano Research Institute Ltd.; Medium-term forecasting of medical bionix (artificial organs) market and thorough analysis of participating companies, 2011, https://www.yano.co.jp/market_reports/C53203100. (in Japanese)

3. Borus, T., and Thornhill, T.; Unicompartmental knee arthroplasty, Journal of the American Academy of Orthopaedic Surgeons, 16(1), pp.9-18, 2008.

4. Nishio, S., Hossain, B., Nii, M., Hiranaka, T., and Kobashi, S.; Workflow recognition from knee surgical videos: role of deep neural networks, The 5th International Symposium on Affective Science and Engineering, 2-B-2, 2019.

5. Twinanda, A. P., Shehata, S., Mutter, D., Marescaux, J., de Mathelin, M., and Padoy, N.; EndoNet: A deep architecture for recognition tasks on laparoscopic videos, IEEE Trans Med Imaging, 36(1), pp.86-97, 2017.

6. Goodfellow, I., Bengio, Y., and Courville, A.; Deep learning, The MIT Press, 2016.

7. Allan, M., et al.; 2017 robotic instrument segmentation challenge, arXiv preprint arXiv:1902.06426, 2019.

8. Iwamoto, Y.; Prosthetic knee arthroplasty towards acquisition of appropriate alignment and balance, Medical View Inc, pp.172-175, 2008. (in Japanese)

9. Hiranaka, T., Nakanishi, Y., Fujishiro, T., Hida, Y., Tsubosaka, M., Shibata, Y., Okimura, K., and Uemoto, H.; The use of smart glasses for surgical video streaming, Surgical Innovation, 24(2), pp151-154, 2017.

10. Jin, Y., Dou, Q., Chen, H., Yu, K., Qin, J., Fu, C. W., and Heng, P.A.; SV-RCNet: Workflow recognition from surgical videos using recurrent convolutional network, IEEE Trans. on Medical Imaging, 37(5), pp.1114-1126, 2018.

11. Namazi, B., Sankaranarayanan, G., and Devarajan, V.; Automatic detection of surgical phases in laparoscopic videos, Proceedings on the International Conference in Artificial Intelligence, pp.124-130, 2018.
12. Hochreiter, S., and Schmidhuber, J.; Long short-term memory. Neural Computation, 9(8), pp.1735-1780, 1997.
13. Simonyan, K., and Zisserman, A.; Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556S, 2014.
14. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H.; Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv:1704.04861, 2017.
15. Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q.; Densely connected convolutional networks, arXiv:1608.06993H, 2016.
16. He, K., Zhang, X., Ren, S., and Sun, J.; Deep residual learning for image recognition, arXiv:1512.03385, 2015.
17. Deng, J., Dong, W., Socher, R., Li, L.-J., Kai, L., and Li, F.-F.; ImageNet: A large-scale hierarchical image database, 2009 IEEE Conf. on Computer Vision and Pattern Recognition, 2009.
18. Kingma, D., and Ba, J.; Adam: A method for stochastic optimization, The International Conference on Learning Representations, 2015.
19. Twinanda, A. P.; Vision-based approaches for surgical activity recognition using laparoscopic and RGBD videos, Université de Strasbourg, 2017.

Shoichi NISHIO (Non-member)
Shoichi Nishio is a graduate student in the graduate school of engineering, department of electronics and information engineering, University of Hyogo. His research interests includes Kansei engineering and Soft Computing approach to Medical Imaging Processing. He is currently working on orthopaedic surgery navigation system using convolutional neural network to intraoperative support operating room nurse.

Belayat HOSSAIN (Non-member)
Belayat Hossain has been serving as a JSPS postdoctoral fellow since Nov 2018 in the graduate school of engineering, University of Hyogo. His research interests include Machine learning in medical image analysis, deep learning, predictive model, computer aided Orthopaedic surgery & biomedical signal analysis. He is also a MEXT scholar and IEEE member.

Manabu NII (Non-member)
Manabu Nii has been serving as associate professor at Graduate School of Engineering, Department of Electronics and Information Engineering in University of Hyogo. His research interests include Soft Computing approach to Medical Imaging Processing.

Naomi YAGI (Non-member)
Naomi Yagi has been serving as associate professor at Himeji Dokkyo University. His research interests include Soft Computing approach to Medical Imaging Processing.

Takafumi HIRANAKA (Non-member)
Takafumi Hirakana has been serving as orthopaedic surgeon at Takatsuki General Hospital. He is an expert in orthopaedic surgery including knee replacement surgery.

Syoji KOBASHI (Non-member)
Syoji Kobashi received BE in 1995, ME in 1997, and Doctor of Engineering in 2000, all from Himeji institute of Technology. He was an assistant professor at Himeji Institute of Technology (2000-2004), an associate professor (2005-2016), currently a professor (2016-) and the manager of advanced medical engineering research center (2016-), University of Hyogo. And, he was a guest associate professor at Osaka University, WPI immunology frontier research center (2010-2016), and was a visiting scholar at University of Pennsylvania (2011-2012). His research interests include medical image understanding and artificial intelligence. He received 16 international awards, including Lifetime Achievement Award (WAC, 2016), Franklin V. Taylor Memorial Award (IEEE-SMCS, 2009). He has been serving on Program Chair of WAC2020, and many others. Moreover, he is an editor-at-large of Intelligent Automation & Soft Computing Journal, an editor-in-chief of International Journal of Biomedical Soft Computing and Human Sciences, etc. He is the senior member of IEEE.