Human-Computer Interaction & Emerging Technologies, Vol. 195, 2025, 230–240
https://doi.org/10.54941/ahfe1006241

Investigation of Image Processing Methods Based on Photographs for Automatic Posture Recognition

Naoki Sugiyama[1], Yoshihiro Kai[2], Hitoshi Koda[3], Toru Morihara[4], and Noriyuki Kida[1]

[1]Kyoto Institute of Technology, Kyoto, Japan
[2]Kyoto Tachibana University, Kyoto, Japan
[3]Kansai University of Welfare Sciences, Osaka, Japan
[4]Marutamachi Rehabilitation Clinic, Kyoto, Japan

ABSTRACT

Japan has the highest aging rate worldwide, underscoring the importance of maintaining daily health for older adults. Postural assessment serves as a valuable indicator of health status. The purpose of this study is to construct an automatic posture recognition model using photographs. As a preliminary investigation, pre-processing methods suitable for machine learning datasets was examined. A total of 278 older adults from sagittal were captured using Kinect v2. the photographs were cropped to exclude non-relevant areas and transformed into grayscale. Subsequently, the cropped images underwent background removed, four edge-detection methods (Prewitt, Sobel, Laplacian 4-neighbors, and Laplacian 8-neighbors), and silhouette extraction, respectively, along with the original images, resulting in seven distinct datasets. A posture the images were classified into Ideal and Non-ideal categories according to physical therapists. The recognition model employed a Support Vector Machine (SVM), with feature extraction methods utilizing Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF). The dataset was divided into training (70%) and test (30%) subsets, with 15 cross-validation sets generated for robustness. Results showed that the Prewitt edge detection method achieved the highest average of F1 score (0.45 ± 0.07) with SIFT, while silhouette extraction yielded the best performance (0.48 ± 0.08) with SURF. The overall accuracy was relatively low; however, when compared to the cropping images, all methods demonstrated higher values, and the order of accuracy was clearly established. These results suggest that further improvements in accuracy could be achieved through tuning the recognition model, highlighting the potential applicability to deep learning frameworks.

Keywords: Posture recognition, Image processing, Sift feature, Surf feature, Support vector machine

INTRODUCTION

Japan faces the world's highest aging rate, making long-term care prevention among the elderly increasingly vital (Cabinet Office Japan, 2022). Care prevention is categorized into three stages: primary prevention to avert disease onset, secondary prevention to address elevated risk stages, and tertiary prevention to inhibit the progression of conditions requiring assistance. Of these, primary and secondary prevention targeting health maintenance are crucial for sustaining and enhancing the quality of life (QOL) in the elderly. Daily muscle strength training is recommended as a key measure in this regard. Within this context, postural assessment has garnered attention as an indicator of muscle strength.

Posture encompasses body position and structure, referring to the relative arrangement of body parts during movement and the unique form of bodily support (Kendall, 2006; Aoki, 2022). Damage to the skin, connective tissues, muscles/fascia, or joints can impair the maintenance of ideal posture, resulting in poor posture. Such postural deterioration, often caused by myofascial imbalances, can induce discomfort and pain. Furthermore, poor posture is linked to both physical and mental health; psychological factors like anxiety and negative emotions have been reported to cause postural changes, such as spinal flexion (Takei, 2013; Oatis, 2012; Maekawa, 2023). Thus, posture serves as a significant indicator of overall health.

Given the increasing older adults population, healthcare professionals face time constraints in conducting comprehensive postural assessments. This situation highlights the need for a screening inspection that allows individuals to easily monitor their health status before seeking clinical consultation. To address this, we explored a photograph-based identification method for automatic posture assessment, emphasizing the importance of appropriate pre-processing to enhance relevant features.

Current visual assessments of photographs struggle to identify detailed postural features, although they can generally classify ideal and poor postures. Moreover, automatic assessment must consider the visual characteristics and information embedded in photographs. Since images often contain irrelevant details such as clothing patterns and backgrounds, using raw photographs is inappropriate. Thus, image processing techniques that highlight target areas while minimizing extraneous information are essential (Shioiri and Omachi, 2011). Previous research has demonstrated the efficacy of silhouette images generated from sagittal-plane photographs, which reduce noise from clothing while preserving appearance outlines (Sugiyama, 2024). These silhouettes achieved comparable accuracy to raw photographs in visual inspection, indicating that unnecessary information can be removed without compromising postural features. Despite this potential, challenges remain in generating silhouette images: semantic segmentation automates the process but compromises region definition accuracy, while manual creation ensures accuracy but incurs higher production costs. Therefore, an automatic and precise processing method distinct from silhouette is necessary.

This study investigates photographs pre-processing methods for automation and evaluates them using machine learning-based discriminative models. We constructed Support Vector Machine (SVM) models using seven image sets: cropping photographs, removed background images, four kinds of edge detection images, and silhouette images. Feature extraction employed Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF).

This study contributes to developing an accessible, automatic posture recognition tool for elderly care prevention. The comparison of SVM-based accuracy across various datasets will identify the appropriate pre-processing techniques. By establishing a robust pre-processing method, we aim to train discriminative models on enhanced datasets, thereby enabling applications such as Convolutional Neural Networks (CNNs).

METHOD

Figure 1 illustrates the workflow employed in this study, detailing the sequence from data acquisition to the construction and evaluation of the identification model.

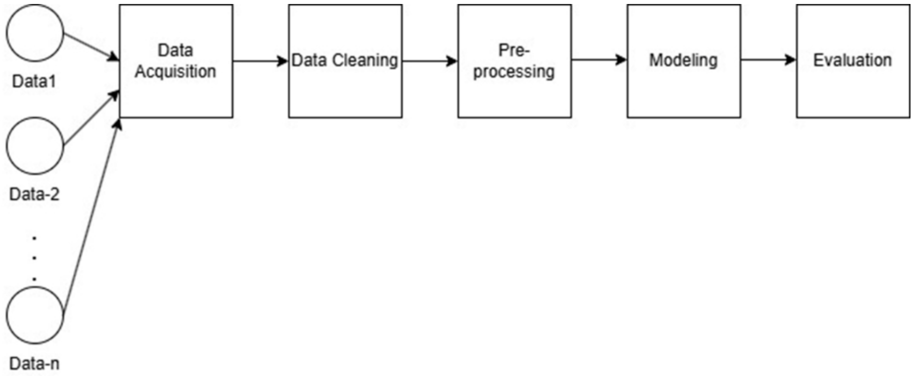

Figure 1: Flow from data acquisition to construction and evaluation of the identification model.

The methodology was structured based on the learning workflow proposed by Zheng et al. (2019), and proceeded through the stages of data acquisition, data cleaning, pre-processing, modeling, and evaluation.

Data Acquisition

The training dataset was collected during older adult physical fitness events conducted in June 2018 and June 2019. Photographs of participants in a stationary standing posture were captured. Participants were instructed to stand on a designated floor mark with their feet shoulder-width apart and arms aligned with their body. Imaging was performed using a Kinect v2 color camera positioned three meters from the right sagittal plane. No specific clothing requirements were imposed during photography. Consequently, a total of 658 images, including duplicates, were obtained.

Posture assessments were conducted by three physical therapists (PTs), each with over 10 years of clinical experience. To ensure the consistency and reliability of these assessments, the criteria were standardized through prior consultation. Postural classifications were based on the Kendall classification, which categorizes posture into four types: Ideal, Kyphosis Lordosis (KL), Sway Back (SB), and Flat Back (FB). These photographs, along with their corresponding posture assessments, comprised the dataset. However, precise classification based solely on visual inspection is challenging (Gadotti, 2013; Sugiyama, 2024), and it is anticipated to be difficult in the image recognition as well. Therefore, posture labels for the photographs were categorized into two groups: ideal and non-ideal (KL, SB, FB). Participation was voluntary, and the study was approved by the Ethics Committee of Kyoto Institute of Technology (protocol number: 2018–19).

Data Cleaning

The photographs were captured using the Kinect v2, simultaneously acquiring depth maps. Depth maps can be utilized to obtain human contours (Sykora, 2014). Considering future methods that combine depth maps and photographs, the initial dataset was unified to include both depth maps and photographs, resulting in a collection of 278 images.

Pre-Processing

Pre-processing aimed to enhance features pertinent to posture identification while minimizing noise. Since the original images contained substantial non-human areas, cropping was initially performed to remove extraneous regions. Subsequently, grayscale conversion was applied to mitigate the influence of clothing colors. Figure 2 shows six pre-processing methods.

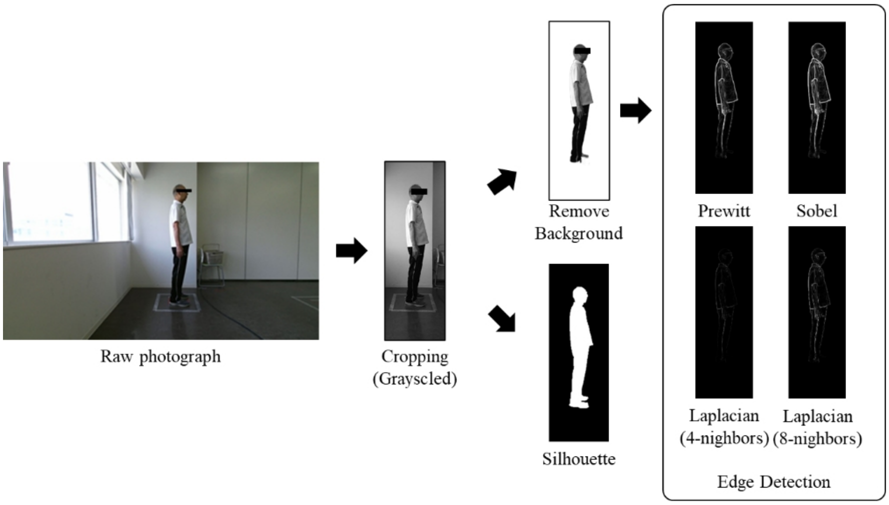

Figure 2: Example of six preprocessing methods applied to a cropped, grayscaled image.

The applied methods included:

- Cropping: Defined by predetermined area specifications to streamline the process.
- Remove Background (Remove BG): Executed using Python’s rembg library (Gatis, 2024), this method further minimized noise, providing a more human-centric dataset compared to simple cropping.

- Edge Detection: Leveraging the abrupt changes in luminance to preserve geometric features while reducing data volume, four edge detection methods were implemented: Prewitt and Sobel (first-order derivatives), and Laplacian 4 and 8 (with 4- and 8-neighbors) configurations (second-order derivatives) (Horikoshi, 2016).
- Silhouette Extraction: Conducted manually using Photoshop's contour extraction function. This process eliminated clothing and background influences, and geometric transformations were applied.

Modeling

A Support Vector Machine (SVM) was adopted for classification, given its suitability for binary classification and robust predictive accuracy for unseen data (Takeda and Toriyama, 2015; Kinjou, 2023). Feature extraction was performed using Scale-Invariant Feature Transform (SIFT) (Lowe, 1999) and Speeded-Up Robust Features (SURF) (Bay, 2006). The dataset was randomly partitioned into a training set (70%) and a test set (30%). To mitigate potential biases from a single data split, 15-fold cross-validation was conducted through repeated random sampling. Model development was used MATLAB (2021R).

Evaluation

The model's discriminative performance was assessed using the test set. Cross-tabulation matrices comparing true and predicted labels were constructed to compute four evaluation metrics: Accuracy, Precision, Recall, and F1 Score, as defined in Equations (1)–(4).

$$\begin{aligned}
\mathit{Accuracy} &= \frac{\mathit{True\ Positive} + \mathit{True\ Negative}}{\mathit{True\ Positive} + \mathit{True\ Negative} + \mathit{False\ Positive} + \mathit{False\ Negative}} && (1)\\
\mathit{Precision} &= \frac{\text{True Positive}}{\mathit{True\ Positive} + \mathit{False\ Positive}} && (2)\\
\mathit{Recall} &= \frac{\text{True Positive}}{\mathit{True\ Positive} + \mathit{False\ Negative}} && (3)\\
\mathit{F1\ Score} &= 2 \times \frac{\mathit{Precision} \times \mathit{Recall}}{\mathit{Precision} + \mathit{Recall}} && (4)
\end{aligned}$$

Each metric provided a comprehensive assessment of model performance. Moreover, the model with the highest accuracy among the 15 iterations was selected for further validation, wherein its predictive validity on unknown data was confirmed via cross-tabulation matrix comparing ground truth and model outputs.

RESULTS

Evaluation of the Constructed Model Using SIFT

Figure 3 illustrates the accuracy obtained from 15-fold cross-validation across the different datasets.

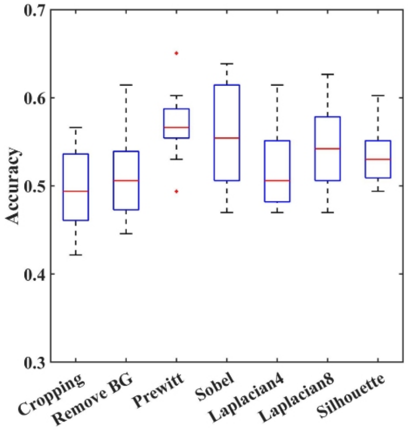

Figure 3: Accuracy for 15-fold cross-validations across datasets.

Among the datasets, the Cropping dataset exhibited the lowest accuracy, with a mean value of 0.49 ± 0.04. Conversely, the highest mean accuracy was observed in the Prewitt dataset, achieving 0.57 ± 0.04. The F1 scores, calculated from the average precision and recall over the 15 iterations, are presented in Figure 4.

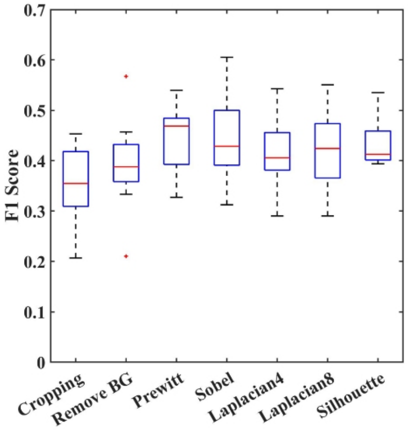

Figure 4: F1 scores for 15-fold cross-validations.

Consistent with the accuracy results, the Cropping dataset yielded the lowest F1 score of 0.35 ± 0.07, while the Prewitt dataset recorded the highest F1 score of 0.45 ± 0.07. Table 1 provides the cross-tabulation matrix corresponding to the model with the best performance among the 15 runs.

Table 1: Cross-tabulation matrix of the identification results for the best-performing model.

		Predicted Label	
		Ideal	**Non-Ideal**
Correct label	Ideal	17	19
	Non-ideal	10	37

For the optimal model using the Prewitt dataset, the accuracy reached 0.65, with an F1 score of 0.54. The sensitivity and specificity were 0.79 and 0.47, respectively.

Evaluation of the Constructed Model Using SURF

Figure 5 presents the accuracy results from 15-fold cross-validation across the datasets.

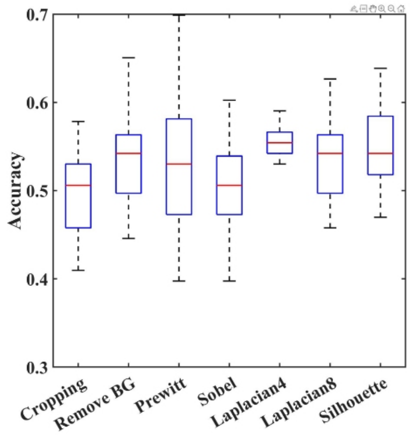

Figure 5: Accuracy for 15-fold cross-validations across datasets.

Similar to the SIFT-based evaluation, the Cropping dataset demonstrated the lowest mean accuracy of 0.50 ± 0.05. In contrast, the Laplacian4 dataset achieved the highest mean accuracy at 0.56 ± 0.02. The F1 scores, averaged over the 15 iterations, are shown in Figure 6.

While the Cropping dataset yielded an F1 score of 0.38 ± 0.08, the Laplacian4 dataset recorded the lowest value at 0.08 ± 0.04. On the other hand, the Silhouette dataset attained the highest F1 score of 0.48 ± 0.08. Although Laplacian4 exhibited the best mean accuracy, the substantial discrepancy in its F1 score indicated a performance bias. Therefore, the Silhouette dataset was prioritized for identifying the model with the highest accuracy. Table 2 shows the cross-tabulation matrix for the best-performing model across the 15 runs.

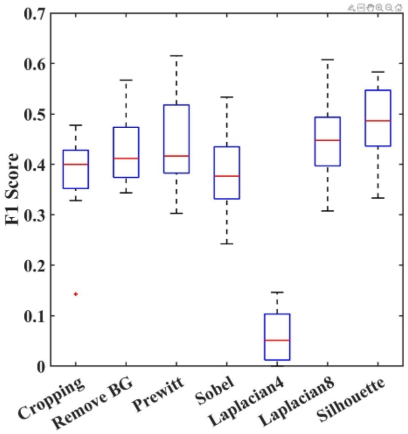

Figure 6: F1 scores for 15-fold cross-validations.

Table 2: Cross-tabulation matrix of the identification results for the best-performing model.

		Predicted Label	
		Ideal	**Non-Ideal**
Correct label	Ideal	21	15
	Non-ideal	15	32

Using the Silhouette dataset, the optimal model achieved an accuracy of 0.64 and an F1 score of 0.58, with sensitivity and specificity values of 0.68 and 0.58, respectively.

DISCUSSION

Model Performance Using SIFT

The application of various treatments across all data sets resulted in higher average values for both Accuracy and F1 Score when compared to untreated Cropping. These findings underscore the importance of noise removal in automatic identification. Among the methods evaluated, Prewitt demonstrated the highest performance metrics. As a fundamental edge detection technique, Prewitt generally exhibits lower edge intensity than Sobel. Nevertheless, certain studies have indicated that Prewitt can more distinctly capture edges in specific images (Baareh, 2018). In this experiment, while Sobel provided enhanced edge detection, it may have also emphasized extraneous features unrelated to posture, such as clothing and hairstyle.

The top-performing model attained an Accuracy of 0.65 and an F1 Score of 0.54, surpassing the criterion threshold of 0.5 and suggesting a marginally favorable trend. The model's sensitivity and specificity were calculated at 0.79 and 0.47, respectively. Given that photographic posture recognition is intended as a screening procedure, prioritizing sensitivity is essential. Consequently, Prewitt emerges as an effective pre-processing technique for photographic data. Although the current identification accuracy remains limited, integrating methods such as SIFT with CNN (Zheng, 2018; Tsourounis, 2022) presents a promising avenue for enhancing the reliability of identification models within the targeted data set.

Model Performance Using SURF

In terms of accuracy, Cropping exhibited the lowest average performance, similar to SIFT. Conversely, the Laplacian4 method achieved the highest Accuracy. However, this tendency did not extend to the F1 Score, where Laplacian4 recorded the lowest value, indicating suboptimal identification pre-processing. Excluding Laplacian4, the Silhouette method attained the highest Accuracy.

The most effective model utilizing Silhouette processing achieved an Accuracy of 0.64 and an F1 Score of 0.58, values comparable to those obtained with Prewitt in conjunction with SIFT. Sensitivity and specificity for this model were 0.68 and 0.58, respectively, reflecting a more balanced performance than that observed with SIFT. According to relevant research, Silhouette has been previously employed in gait identification research (Han, J., & Bhanu, B., 2006), supporting its applicability to static standing posture recognition in this study. While Silhouette is suggested to be a valid pre-processing option, the overall discrimination accuracy remains insufficient. Nonetheless, further accuracy improvements are anticipated through the integration of methods like SURF with CNN (Elmoogy, 2020) and the continued application of deep learning techniques.

CONCLUSION

This study explored photograph pre-processing methods for automation and evaluated the performance of discriminative models based on SIFT and SURF using Support Vector Machines (SVM). The key findings are summarized as follows.

For the SIFT-based SVM, the Prewitt dataset yielded the highest classification accuracy. The optimal model achieved an accuracy of 0.65 and an F1 score of 0.54, with sensitivity and specificity values of 0.79 and 0.47, respectively. Although the overall accuracy was not sufficiently high, the clear trend in learning accuracy across different pre-processing methods indicates the potential applicability of the Prewitt in deep learning contexts.

In the case of the SURF-based SVM, the Silhouette process demonstrated relatively high accuracy with reduced bias. The best model recorded an accuracy of 0.64 and an F1 score of 0.58, comparable to the results obtained with the Prewitt dataset in the SIFT model. The sensitivity and specificity values were 0.68 and 0.58, respectively, reflecting less bias than the Prewitt. While the discrimination accuracy remains moderate, the findings suggest that the Silhouette process is a viable pre-processing step for automatic posture recognition.

Overall, this study provides a quantitative evaluation of the impact of various pre-processing methods on classification accuracy for automatic posture recognition. The results offer valuable guidelines for selecting appropriate pre-processing techniques. Future work will focus on integrating these methods with advanced deep learning-based identification models to enhance accuracy for practical applications.

INSTITUTIONAL REVIEW BOARD STATEMENT

The Ethics Committee of the Kyoto Institute of Technology approved this study (Protocol Number 2018-19—22 June 2018).

ACKNOWLEDGMENT

We appreciate all the research staff and PTs who supported our research and all the individuals who corroborated the contents of this manuscript.

REFERENCES

Aoki, T., & Hayashi, N. (2022). Functional Anatomical Palpation Technique for Exercise Therapy - Upper Extremity, 2nd Edition with Video: Medical View Co., Ltd. p. 388 ISBN: 978-4-7583-2093-1.

Baareh, A. K. M., Al-Jarrah, A., Smadi, A. M., & Shakah, G. H. (2018). Performance Evaluation of Edge Detection Using Sobel, Homogeneity and Prewitt Algorithms. Journal of Software Engineering and Applications, 11(11), pp. 537–551.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded Up Robust Features. European Conference on Computer Vision, pp. 404–417.

Cabinet Office Japan. Annual Report on the Ageing Society. 2022. https://www8.cao.go.jp/kourei/english/annualreport/2022/pdf/2022.pdf

Elmoogy, A. M., Dong, X., Lu, T., Westendorp, R., & Tarimala, K. R. (2020). SurfCNN: A Descriptor Accelerated Convolutional Neural Network for Image-Based Indoor Localization. IEEE Access, 8, pp. 59750–59759.

Gadotti, I. C., Armijo-Olivo, S., Silveira, A., & Magee, D. (2013). Reliability of the Craniocervical Posture Assessment: Visual and Angular Measurements Using Photographs and Radiographs. Journal of Manipulative and Physiological Therapeutics, 36(9), pp. 619–625.

Gatis, D. (2024). rembg. Available at: https://github.com/danielgatis/rembg.

Han, J., & Bhanu, B. (2006). Individual Recognition Using Gait Energy Image. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(2), pp. 316–322.

Horikoshi, R., Morimoto, M., Miura, Y., & Sawano, H. (2016). Image Engineering. [Translated from Japanese]: Ohmsha, Ltd. p. 232. ISBN: 978-4-274-22007-4.

Kendall, F. P., McCreary, E. K., and Provance, P. G. (2005). Muscles: Testing and Function, with Posture and Pain. Philadelphia: Lippincott Williams and Wilkins. p. 560. ISBN: 10–0781747805.

Kinjou, T. (2023). Understanding and Applying Machine Learning Techniques and Mechanisms for Work [Translated from Japanese]: SHUWA SYSTEM CO., LTD. p. 304. ISBN: 978-4-7980-6687-5.

Lowe, D. G. (1999). Object Recognition from Local Scale-Invariant Features. Proceedings of the International Conference on Computer Vision, pp. 1150–1157.

Maekawa, M., Yoshizawa, E., Hayata, G., & Ohashi, S. (2023). Physical and psychological effects of postural educational intervention for students experienced school refusal. Current Psychology, 42, pp. 3510–3519.

Oatis, C. A., Yamazaki, A., Sato, S., Shirahoshi, S., Fujikawa, T., & Ikeya, M. (2012). Kinesiology: the mechanics and pathomechanics of human movement: Round Flat, Inc. ISBN-10: 4904613198.

Shioiri, R., & Omachi, S. (2011). Image Information Processing Engineering [Translated from Japanese]: Asakura Publishing Co., Ltd.. p. 144. ISBN: 978-4-254-22888-5.

Sugiyama, N., Kai, Y., Koda, H., Morihara, T., & Kida, N. (2024). Agreement in the Postural Assessment of Older Adults by Physical Therapists Using Clinical and Imaging Methods. Geriatrics, 9, 40.

Sykora, P., Kamencay, P., & Hudec, R. (2014). Comparison of SIFT and SURF Methods for Use on Hand Gesture Recognition based on Depth Map. AASRI Procedia, 9, pp. 19–24.

Takeda, I., & Toriyama, M. (2015). Support Vector Machines [Translated from Japanese]: Kodansha Ltd. p.192. ISBN: 978-4-06-152906-9.

Takei, H. (2013). Evaluation and treatment for posture. Spinal Surgery, 27(2), pp. 119–124.

Tsourounis, D., Kastaniotis, D., Theoharatos, C., Kazantzidis, A., & Economou, G. (2022). SIFT-CNN: When Convolutional Neural Networks Meet Dense SIFT Descriptors for Image and Sequence Classification. J. Imaging, 8, 256.

Zheng, A., Casari, A., & Hokusoemu Co., Ltd. (Trans.). (2019). Feature Engineering for Machine Learning: Principles and Practice with Python [Translated from Japanese]: O'Reilly Japan. p. 224. ISBN-10: 4873118689.

Zheng, L., Yang, Y., & Tian, Q. (2018). SIFT Meets CNN: A Decade Survey of Instance Retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40, pp. 1224–1244.