Journal of KIIT. Vol. 23, No. 9, pp. 153-160, Sep. 30, 2025. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2025.23.9.153

Shadow Removal and Image Segmentation for Grayscale Hand Gesture Images

Jong Gwan Lim*

Abstract

While numerous shadow removal algorithms have been proposed for the increasingly prevalent color images, they often suffer from degraded performance in grayscale images. This paper presents a novel image segmentation method for detecting and removing shadows to enable hand gesture recognition in low-quality grayscale images. By applying Mean Shift clustering, the proposed method achieves efficient processing and distinguishes shadow and non-shadow regions at the cluster level using the geometric features of the hand and local minima near its contours. In the final stage, cluster-level penumbra regions overlapping with the hand contours are identified and removed from the detected hand region clusters to improve pixel-wise accuracy. Through both quantitative and qualitative analyses, the effectiveness of the method is validated, and it is confirmed that distortion-free segmentation is achieved even in shadow-free images.

요 약

범람하는 컬러 영상과 이를 위한 다수의 그림자 제거 알고리즘이 소개되고 있으나 회색조 영상에서 성능이 저하한다. 본 논문은 저품질 회색조 영상에서 손 제스처 인식을 위해 그림자를 검출하고 제거하는 새로운 영상 분할 방법을 제안한다. Mean Shift 클러스터링을 도입해 효율적인 처리를 가능하게 하고, 손의 기하학적인 특징과 윤곽선 근처에서 발생하는 국지 최저값에 근거해 그림자/ 비그림자 영역을 클러스터 단위로 구분한다. 마지막 단계에서, 손의 윤곽선과 겹쳐 발생하는 클러스터 단위 그림자 영역을 검출하고 제거함으로써 화소 단위 정확성을 개선한다. 정략적인 분석과 정성적인 분석을 통해 성능의 우수성을 검증하고 그림자가 없는 영상에서도 왜곡없는 영상 분할을 달성함을 확인한다.

Keywords
segmentation, shadow removal, hand gesture, grayscale, meanshift

* Professor, Dept. of Robotics, Mokwon Univ.
- ORCID: https://orcid.org/0000-0003-1223-0279

· Received: Aug. 13, 2025, Revised: Sep. 08, 2025, Accepted: Sep. 11, 2025
· Corresponding Author: Jong Gwan Lim
Dept. of Robotics, Mokwon University, 88, Doanbuk-ro, Seo-gu, Daejeon, Korea
Tel.: +82-42-829-7520, Email: jonggwanlim@gmail.com

Ⅰ. Introduction

Hand gestures convey diverse meanings shaped by cultural and situational contexts, serving as key modalities in human communication, affective expression, and command transmission. Vision-based gesture recognition technologies are integral to contactless interfaces, necessitating precise extraction of finger position and morphology [1]. Binarization, a fundamental preprocessing step, reduces image complexity and enhances the robustness of hand shape recognition by isolating hand regions from the background [2].

However, shadows present a significant challenge in image processing tasks—including object recognition, feature extraction, and scene interpretation—by degrading binarization accuracy. In single-frame imagery, shadows obscure object boundaries, distort contours, and diminish brightness, leading to misclassification or omission of hand regions. Additionally, shadowed areas often exhibit color and texture similarities with the background, causing segmentation failures such as object-shadow merging or boundary misidentification. These limitations, documented in prior studies [2][3], underscore the need for effective shadow detection and removal to improve binarization fidelity.

Ⅱ. Background

2.1 Understanding shadows

A shadow comprises distinct regions based on light obstruction: the umbra, where the light source is fully occluded, yields a sharp and high-contrast shadow; the penumbra, formed by partial occlusion, produces softer shadows with blurred edges and reduced luminance, complicating boundary detection in image analysis. The antumbra, occurring when the occluding object is smaller than the light source, is excluded from this study as it is not relevant to the current context (see Fig. 1).

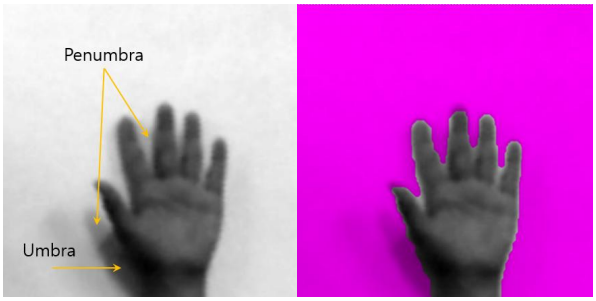


Fig. 1. Type of shadows and ground-truth label

2.2 Related research

Brightness-based thresholding is the most widely adopted method for shadow detection and removal, exploiting the lower pixel intensity of shadow regions relative to their surroundings [3]. Common implementations utilize 1D histograms, with enhancements via 2D histograms [4] and adaptive thresholding. While suitable for real-time applications, this approach struggles with hand images where umbra regions overlap object contours (see Fig. 1).

Illumination modeling, a physics-based technique, estimates lighting conditions to isolate shadows. The observed intensity, I, is modeled as the product of reflectance, R, and illumination, L, and transformed logarithmically into a linear additive form. Reflectance, R, invariant to lighting, is then extracted for shadow detection [5]. Despite its theoretical rigor, this method requires complex estimation and often lacks robustness.

Geometric approaches, based on object-shadow projection rules, are limited by their dependence on object-specific geometry and are unsuitable for hand gesture recognition [6]. Texture-based segmentation has also been explored [7], but performs poorly when texture is weak or ambiguous.

Recent advances in deep learning and machine learning [8] offer promising alternatives, yet most models are optimized for high-resolution color images with three channels. In grayscale contexts, its

performance degrades significantly. For example, [9] applies SLIC (Simple Linear Iterative Clustering), a superpixel algorithm based on K-means, which computes color and spatial distances in Lab space— unsuitable for single-channel grayscale images. Comparative results for each method are shown in Fig. 2.

This paper proposes a novel methodology that addresses the shortcomings of previous studies for grayscale hand gesture images. Instead of relying on inaccurate illumination modeling, the proposed method estimates the direction of the light source to eliminate its influence. In place of the SLIC algorithm, which is unsuitable for grayscale images, Mean Shift clustering is introduced for compatibility with single-channel data. By incorporating the geometric characteristics of the hand and leveraging the observation that contour regions within shadows exhibit minimum brightness, the method selectively identifies shadow and non-shadow Mean Shift clusters.

III. Proposed Method

3.1 preprocessing

As a basic preprocessing step, Gaussian blur or median blur is applied. Then, to estimate the direction of light, portions from the four corners of the image are extracted, and the brightness gradient is estimated using the least squares method. This gradient is then removed from the original image to normalize the background illumination. Then, to maximize contrast within the image, gamma correction (Eq. 1) is applied twice ($\gamma < 1$, $\gamma > 1$). Afterward, basic binarization is performed using the Otsu method, followed by erosion operations to extract contours, edge(x,y).

$$I_{corrected} = I^{\gamma} \tag{1}$$

3.2 Mean shift

Mean Shift is an unsupervised clustering algorithm that distinguishes itself from K-Means by not requiring a predefined number of clusters. The algorithm begins by scattering multiple data points across the image and performing kernel density estimation around each point. Each point is iteratively shifted toward regions of higher density to identify cluster centers.

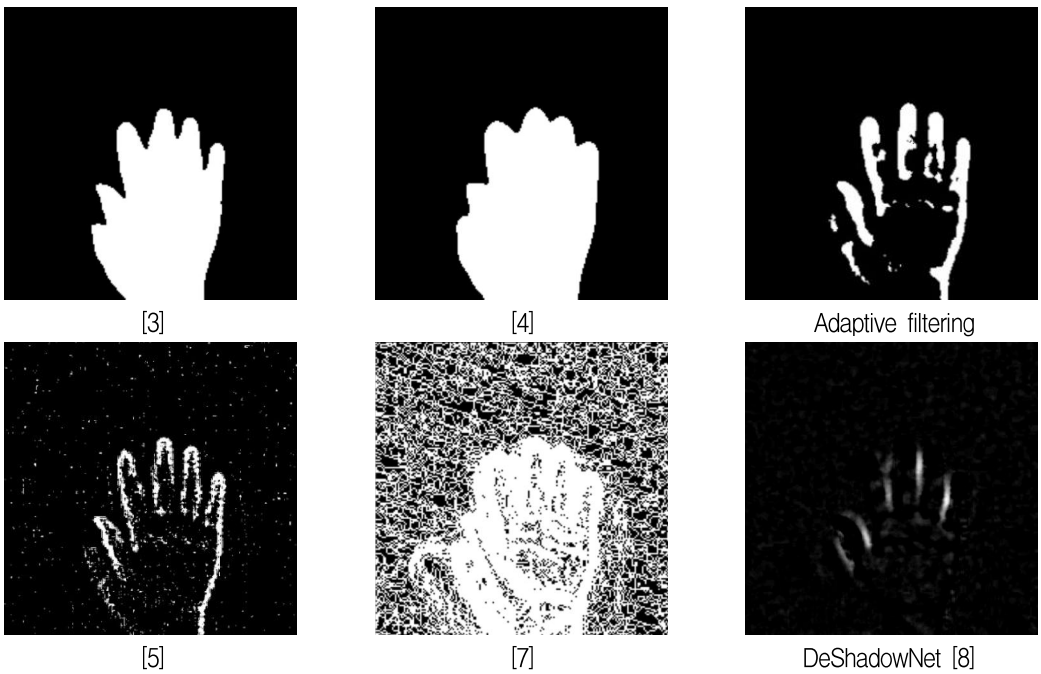


Fig. 2. Related research results

A Gaussian kernel is typically employed, with its size determined by a bandwidth parameter; a larger bandwidth results in fewer clusters due to broader kernel coverage. The areas where data points converge are classified into clusters, and each cluster is assigned the average brightness value of the pixels it contains.

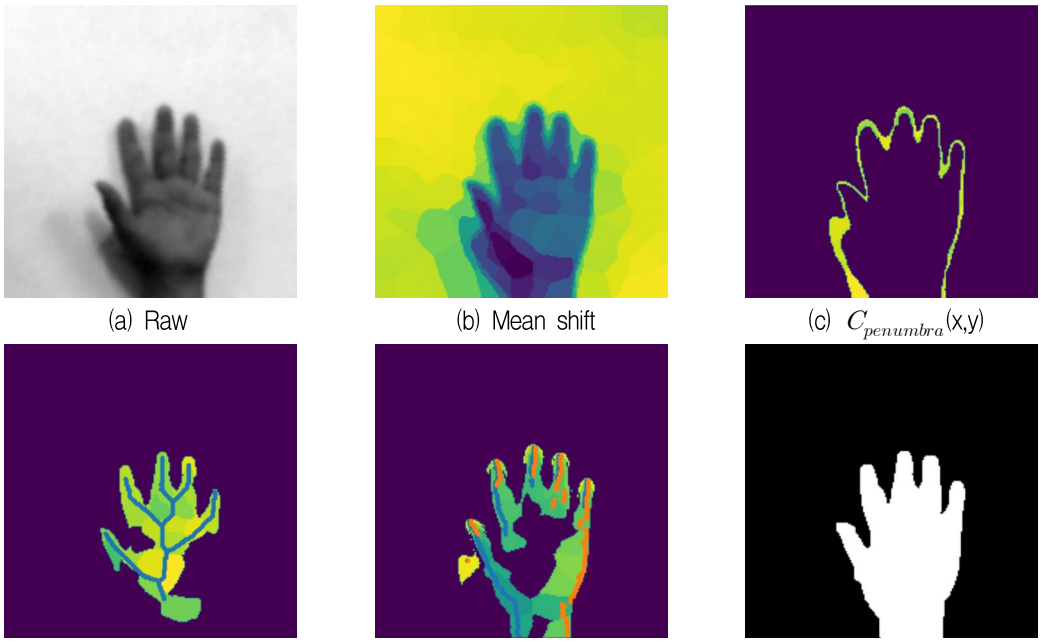
Mean Shift is widely used in image segmentation and offers several advantages. It groups image pixels into superpixels, enhancing computational efficiency, and maintains spatial consistency and object contours at the cluster level. Furthermore, it requires only a few parameters, making parameter tuning straightforward. In the present task, which involves grayscale images, clustering can be performed solely based on brightness values without relying on color channel information (see Fig. 3(b))

3.3 Penumbra detection

As shown in Fig. 1, penumbra tend to appear faintly near the contours of the hand. Based on this observation, the proposed method detects Mean Shift clusters that intersect with the previously extracted contours, denoted as edge(x, y). Since edge(x,y) is derived from a binarized image where shadow removal may be incomplete, it can be inaccurate. To address this, the shape of the clusters overlapping with the contours is used as a distinguishing factor. Most clusters corresponding to penumbra are thin and elongated either horizontally or vertically. Therefore, the ratio between the number of pixels where edge(x, y) overlaps and the total number of pixels in each cluster is used to differentiate them, serving as the second parameter of the proposed methodology. The clusters identified as penumbra regions are denoted as $C_{penumbra}$(x,y) (see Fig. 3(c)).

3.4 Skeletonization

Despite belonging to the same hand region, brightness may vary depending on the shape and curvature of the fingers as well as the position of the light source. Consequently, some Mean Shift clusters are brightness outliers, making uniform thresholding infeasible.



(a) Raw (b) Mean shift (c) $C_{penumbra}$(x,y)

(d) C_{skel}(x,y) (e) C_{hand}(x,y) (f) Final

Fig. 3. Process

To address this issue, a skeletonization process is applied to extract the structural skeleton from the basic binarized image generated during preprocessing [10]. Skeletonization reduces the binary image to its central lines while preserving the overall shape, thereby providing a compact representation of structural features. Clusters intersected by the skeleton, denoted as C_{skel}(x,y), are then selected (see Fig. 3(d))

3.5 Hand region detection

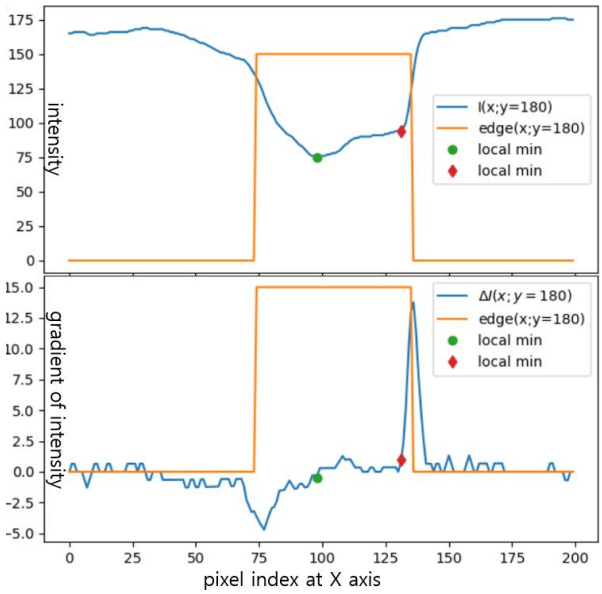


Fig. 4. Local minimum detection

In the case of the umbra region, shadows tend to be darker and are often coupled with the curvature of the hand, making boundary detection particularly challenging. As shown in Fig. 3(b), the brightness of the umbra region is similar to that of clusters within the hand, making pixel-level operations insufficient for distinguishing them in grayscale images. To overcome this limitation, we propose a method that extracts local minima near the contours. Specifically, the brightness gradient along the y-axis, denoted as ΔI(x,y), is computed from the intensity I(x, y), and points where ΔI(x, y) = 0 are identified to locate local extrema closest to the contour. Fig. 4 illustrates the brightness values and their derivatives at y = 180. Compared to conventional methods that detect contours at gradient extrema, our approach reveals that actual boundary values tend to appear further inward, highlighting the limitations of existing techniques. Clusters intersecting with these extracted local extrema, denoted as C_{hand}(x,y), are selected, and their contours are defined as the boundary (see Fig. 3(e)).

3.6 Postprocessing

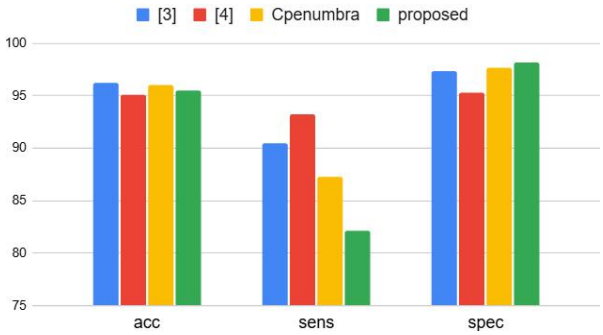
$$(C_{skel} \cup C_{hand}) - C_{penumbra} \tag{2}$$

The three clusters previously selected are combined as shown in Eq. 2 to extract the entire hand region. Since some clusters may be missed by C_{skel}(x,y) and C_{hand}(x,y) within the hand area, or shadow-related clusters may remain outside the hand due to omissions by $C_{penumbra}$(x,y), connected component analysis is processed to refine the result (see Fig. 3(f))

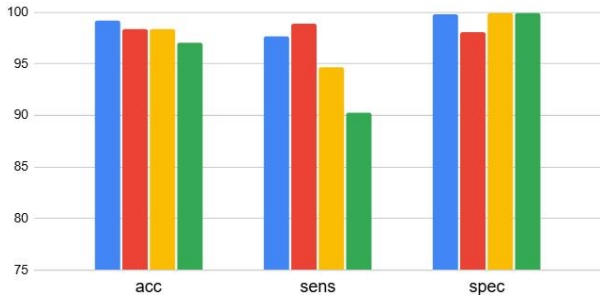
IV. Performace Evaluation

For performance evaluation, we utilize the dataset used in previous studies [2][3]. It consists of 2,000 grayscale images of Chinese numeral hand gestures, captured from 17 subjects across 10 gesture classes, with 200 images per class. The images were recorded using a webcam at a resolution of 200×200 pixels, exhibiting variations in hand size, orientation, lighting conditions, and wrist exposure. From this dataset, 134 images that failed binarization due to shadows and 126 shadow-free images were selected and categorized into a shadowed group and a clean group, respectively. This categorization enables assessment of the impact of the shadow removal methodology on images unaffected by shadows. For performance metrics, we employ pixel-wise accuracy, sensitivity, and specificity. Sensitivity reflects the accuracy of hand region detection, while specificity indicates the accuracy of shadow region detection. To compute

these metrics, pixel-wise labeling was manually performed for each image.



(a) Shadowed group



(b) Clean group

Fig. 5. Quantitative performance evaluation

In the quantitative performance analysis shown in Fig. 5, we compare the accuracy, sensitivity, and specificity of the proposed method with those of the 1D Otsu method [3], the 2D Otsu method [4], and the result from [3] with $C_{penumbra}$(x,y) excluded. Methods [5], [7], and [8] are excluded due to their failure to delineate a closed contour around the hand region. As shown in Fig. 5, the upper row presents results for the shadowed group, while the lower row corresponds to the clean group. The most notable variation occurs in sensitivity: the proposed method exhibits an 8.3% decrease compared to [3] and a 5.15% decrease relative to the penumbra-removed baseline. This reduction is primarily attributed to two factors. First, as illustrated in Fig. 1, the ground truth labels slightly exceed the actual hand boundaries, introducing teacher noise. This issue is exacerbated in shadow-contaminated images, where manual labeling becomes visually ambiguous. Second, the proposed method tends to undersegment the hand region, resulting in a smaller detected area than the ground truth. To account for this, a clean group test is included in the analysis.

In terms of specificity, the proposed method improves by 0.88% over [3] and 2.89% over [4] in the shadowed group, while yielding comparable results to [3] in the clean group—closely aligning with ground truth. These findings suggest that the proposed method effectively suppresses shadows in contaminated images, while maintaining binarization integrity in clean conditions. The modest gain in specificity is largely due to the low proportion of shadow pixels in the overall image. Furthermore, the method employs Mean Shift clustering for shadow removal, functioning as a low-pass filter. Consequently, performance is influenced by the bandwidth parameter and the intersection ratio between hand contours and outer clusters. In this study, the bandwidth is set to 0.005 and the intersection ratio to 15%.

Significant performance improvements are more evident in qualitative comparisons than in quantitative results. We compared outcomes across 260 cases, with a subset illustrated in Fig. 6. Fig. 6(a) - (d) represent samples from the shadowed group, while (e) and (f) show examples from the clean group. The results of the proposed in Fig. 6 reveal that the lower sensitivity observed in the quantitative analysis is mainly due to rough boundaries caused by cluster-based segmentation and the tendency of the proposed method to produce slightly contracted hand regions compared to the ground truth. However, these characteristics do not critically hinder successful shadow removal. Through qualitative analysis, we confirm that even small differences in quantitative metrics are meaningful and should not be overlooked.

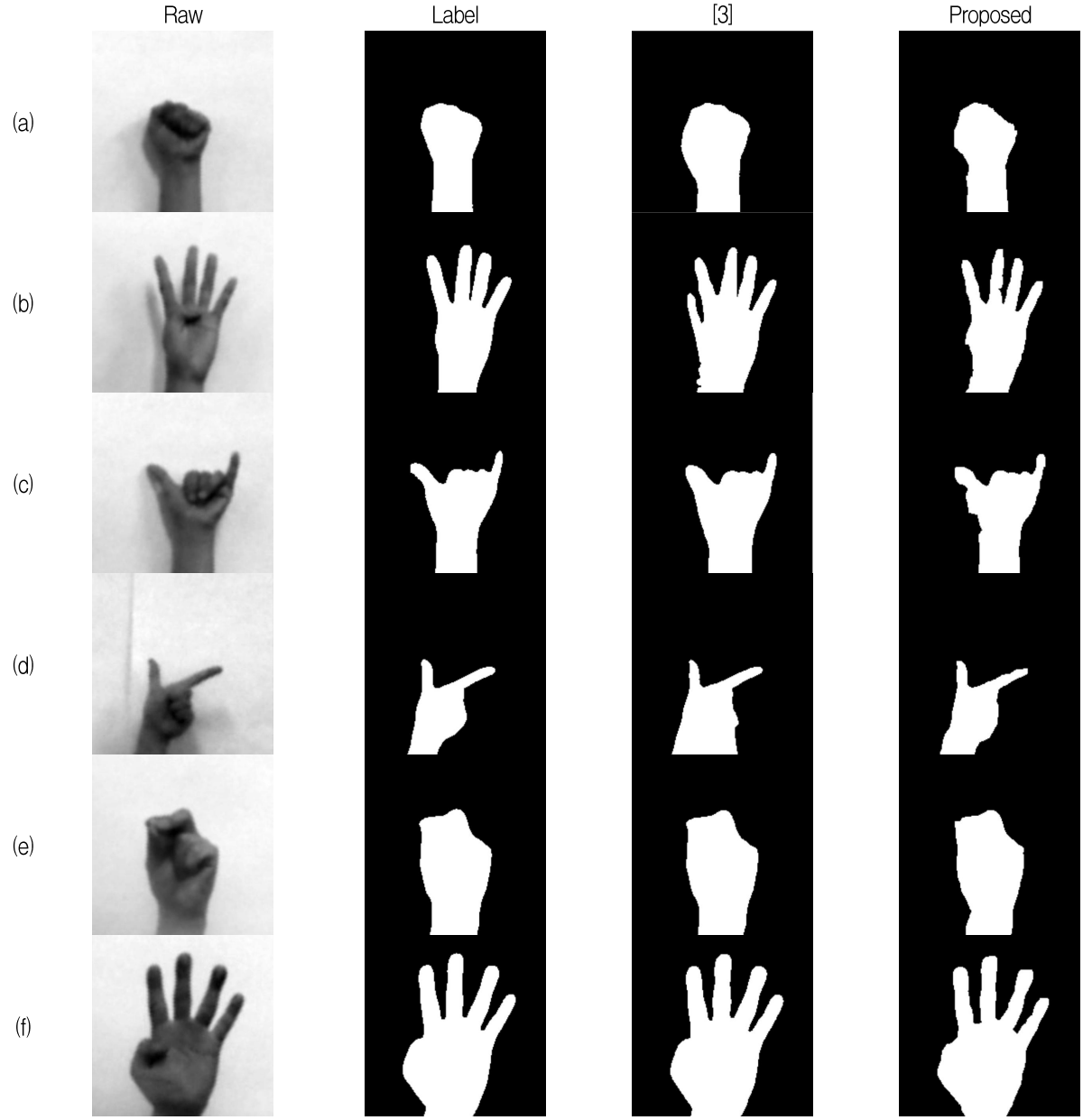


Fig. 6. Qualitative performance evaluation

Ⅴ. Conclusion

This study addresses the critical challenge of shadow detection and removal in low-quality grayscale hand gesture images, which is a task that significantly impacts the accuracy of image segmentation. While numerous algorithms have been developed for color image processing, their effectiveness in grayscale contexts remains limited due to the absence of chromatic information and the reliance solely on brightness values. Factors such as varying illumination, gesture-induced brightness fluctuations, and the natural curvature of the hand further complicate shadow removal in grayscale imagery.

To overcome these limitations, we propose a novel methodology that transforms pixel-level data into cluster-level representations using the Mean Shift. Clusters corresponding to the hand region are identified based on skeletal structure and local brightness minima near contour boundaries. Penumbra

are subsequently removed through contour-based analysis, enabling robust binarization. Although the proposed method tends to slightly contract the outer boundary of the palm region during the removal of umbra and penumbra, it achieves satisfactory performance in both shadow detection and removal. Moreover, the method demonstrates consistent applicability to shadow-free images without compromising segmentation quality.

It is important to note that the performance of the proposed approach is sensitive to two key parameters: the bandwidth of the Mean Shift algorithm and the intersection ratio between contours and peripheral clusters used for penumbra detection. Future research will focus on systematically optimizing these parameters to further enhance the robustness and generalizability of the method.

References

[1] J. QI, L. Ma, Z. Cui, and Y. Yu, "Computer vision-based hand gesture recognition for human-robot interaction: a review", Complex & Intelligent Systems, Vol. 10, pp. 1581-1606, Oct. 2024. https://doi.org/10.1007/s40747-023-01173-6.

[2] J. G. Lim, "Chinese Hand Number Gesture Recognition by Enhanced Multi-stage Template Matching", Journal of KIIT, Vol. 18, No. 8, pp. 115-121, Aug. 2019. http://dx.doi.org/10.14801/jkiit.2019.17.8.115.

[3] J. G. Lim, "Multi-stage Template Matching for One Hand Numerical Gesture Recognition", Journal of KIIT, Vol. 16, No. 5, pp. 15-21, May 2018. https://doi.org/10.14801/jkiit.2018.16.5.15.

[4] C. Sha, J. Hou, and H. Cui, "A robust 2D Otsu's thresholding method in image segmentation", Journal of Visual Communication and Image Representation, Vol. 41, pp. 339-351, Nov. 2016. https://doi.org/10.1016/j.jvcir.2016.10.013.

[5] K. H. Park, "Linear Regression-based 1D Invariant Image for Shadow Detection and Removal in Single Natural Image", Journal of Digital Contents Society, Vol. 19, No. 9, pp. 1787-1793, Sep. 2018. https://doi.org/10.9728/dcs.2018.19.9.1787.

[6] Y. H. Jang, et al., "Design and Implementation of Image Detection System Using Vertical Histogram - Based Shadow Removal Algorithm", Journal of the Korea Institute of Information and Communication Engineering, Vol. 24, No. 1, pp. 91-99, Jan. 2020. https://doi.org/10.6109/jkiice.2020.24.1.91.

[7] X. Yi and M. Eramian, "LBP-Based Segmentation of Defocus Blur", IEEE Transactions on Image Processing, Vol. 25, No. 4, pp. 1626-1638, Apr. 2016. https://doi.org/10.1109/TIP.2016.2528042.

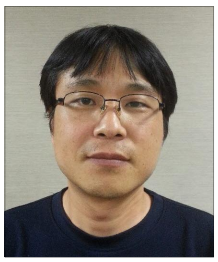
[8] L. Guo, et al., "Single-image shadow removal using deep learning: A comprehensive survey", arXiv preprint, arXiv:2407.08865, Jul. 2024. https://doi.org/10.48550/arXiv.2407.08865.

[9] C. G. Woo, Y. H. Kim, and K. H. Park, "Effective Shadow Region Detection Method using Clustering Algorithms", Journal of Digital Contents Society, Vol. 21, No. 1, pp. 251-257, Jan. 2020. https://doi.org/10.9728/dcs.2018.19.9.1787.

[10] T. C. Lee, R. L. Kashyap, and C. N. Chu, "Building Skeleton Models via 3-D Medial Surface Axis Thinning Algorithms", CVGIP: Graphical Models and Image Processing, Vol. 56, No. 6, pp. 462-478, Nov. 1994. https://doi.org/10.1006/cgip.1994.1042.

Authors

Jong Gwan Lim



2016. 2 : PhD, ME, KAIST
2017. 3 ~ Present : Professor, Dept. of Robotics, Mokwon Univ.
Research interest : HRI, image/signal processing, machine learning