Image Enhancement for Hand Sign Detection

Jing-Wein Wang[1], Tzu-Hsiung Chen[2] and Tsong-Yi Chen[3]
[1]Institute of Photonics and Communications, National kohsiung University of Applied Sciences, Kaohsing, Taiwan
[2] Computer Science and Information Engineering, Taipei Chengshih University of Science and Technology, Taipei, Taiwan
[3]Electronic Department, National kohsiung University of Applied Sciences, Kaohsing, Taiwan

Keywords: Compact Hand Extraction, Singular Value Decomposition Based Image Enhancement, Illumination Compensation.

Abstract: This paper proposes compact hand extraction to assist in computerized handshape recognition. We devised an image enhancement technique based on singular value decomposition to remove dark backgrounds by reserving the skin color pixels of a hand image. The polynomial approximation YCbCr color model was then used to extract the hand. After alignment, we applied illumination compensation to the adaptable singular value decomposition. Experimental results for images from our database showed that our method functioned more efficiently than conventional ones that do not use compact hand extraction against complex scenes.

1 INTRODUCTION

Handshape is an active area of research in visual studies, mainly for handshape recognition and human computer interaction (HCI). The goal of handshape interpretation is to advance human-machine communication so that it resembles human-human interactions more closely. Handshape recognition in an image poses a challenge because such recognition must locate a hand with no prior knowledge regarding its scale, location, pose, and image content. Background and illumination are also problems not yet fully resolved, and numerous other factors can contribute to the external variability of in-plane and out-of-plane rotations. Over the last decade, several methods of applications in advanced handshape interfaces for HCI have been suggested, but these differ from one another in their models. Some of these models are referred to in the current research (Farouk et al.; 2009, Thangali et al., 2011, Holt et al., 2009)

To detect a hand from an image, the whole image is scanned exhaustively to find the likely area of the hand pattern, and then a location and boundary description of that area is created. The initial screening scheme is critical and can reduce the subsequent time spent on processing; however, poorly performed segmentation may disfigure the image of the hand. A common strategy for hand detection is skin-based matching, which determines the image pixels that could represent a shade of human skin (Murthy et al., 2009, Butalia et al., 2010, Khan et al., 2008, Rehrl et al., 2010). This approach provides robustness and automation for holistic descriptions, and serves as a front end for hand extraction from a complex background. An example of the skin-based matching approach is the color modeling approach (Kim et al., 2008) applied in the hue-saturation-intensity (HSI) color space. This model was built by adopting B-spline curve fitting to devise a mathematical model for describing the statistical characteristics of skin color with respect to intensity. Although the color segmentation method based on B-spline curve fitting has been shown to be a powerful learning algorithm for skin color detection, the method of fitting four-bar graphs to continuous curves relies mostly on the quality and quantity of the training data. The uniform color space defined by the International Commission on Illumination (CIE) is known as L*a*b*, which has a more compact skin color cluster than RGB or HSI color spaces (Yin and Xie, 2007). To optimize the use of limited training data, a Restricted Coulomb Energy (RCE) neural network was designed to represent the L*a*b* color values of a pixel, wherein the middle-layer cells embed information on skin color, and the output layer communicates with the corresponding color class. Although the RCE neural network can classify the input color

Wang J., Chen T. and Chen T..
Image Enhancement for Hand Sign Detection.
DOI: 10.5220/0005059101860192
In *Proceedings of the 11th International Conference on Signal Processing and Multimedia Applications* (SIGMAP-2014), pages 186-192
ISBN: 978-989-758-046-8

signal as a skin color class and identify the pixels represented by this color signal as skin texture, some non-skin pixels can be falsely detected because of lighting conditions. Other approaches use a background-subtraction strategy (Jimenez-Hernandez, 2010), wherein a component image can be segmented easily. However, this method works only for particular conditions related to the speed of objects and the frame rate, and is highly sensitive to the frame difference threshold. Adan et al. (2008) presented a hand biometric system for verification and recognition purposes that relied on the natural reference system (NRS) based on a natural hand layout. Although neither hand pose training nor a prefixed position is required in the registration process, users have to extend their hand fully. For successful recognition to occur, only a small degree of rotation is allowed, and the background must be fixed initially.

Although skin color (Kakumanu et al., 2007) differs across ethnic groups, they are distributed in a narrow range on the chrominance plane. Variability in skin tone under varying illumination conditions results from different intensities. Because the color-based method can encounter problems in detecting skin color robustly against a complex background, a potential strategy is to start with low-level cues corresponding to the early attentive visual features in biological vision, such as the palm and fingers (Kumar et al., 2003), and to combine these by anthropometry for locating a potential target for hand detection. Handshapes can also be described by skeleton analysis (Bakina, 2011), which allows a comparison of hands with separated fingers as well as with closed fingers. Nevertheless, this method is restricted to a simple background in a fixed manner as presented in Adan et al. (2008).

A single detector may not cope effectively with variations in the foreground hand while simultaneously discriminating between the foreground and background, especially in applications in which lighting varies widely. To avoid explicit detection of the foreground hand object, we propose an enhancement-before-detection strategy. Contrast enhancement can significantly improve the discovery of an image by emphasizing the background and removing dark objects before performing skin color segmentation. This alleviates the burden of skin-color modeling and focuses instead on labeling only the true skin pixels. A novel aspect of the proposed method involves exploiting the possible combinations of true skin color detection and contrast enhancement for robust hand extraction. The contrast method relies on differences in depth of focus between the foreground hand object and the background environment. An accurate and efficient method for hand extraction is still lacking for color images with a cluttered background, illumination and posture alterations, in-plane and out-of–plane rotations, and scale variations because such conditions complicate the detection of hand features. Years of experimental research have shown that each type of detection technique performs better for detecting isolated features. Therefore, for every selected feature, a fusion of methods from both categories should provide more stable results than one method alone. Based on this reason, and motivated by the observation of the "paw" shape of human hands, we propose complementary techniques based on contrast enhancement and skin color detection. The goal of our approach is to provide an efficient system that operates on complex backgrounds and tolerates illumination and scale variations, and moderate rotations of up to approximately 15˚.

In the next section of this paper, we describe the compact hand-extraction problem, including cascade processes in background removal, disassembling and recombining fingers based on hand anatomy, and illumination compensation. Section 3 presents a discussion of the experimental result with and without compact hand extraction, to corroborate the proposed framework. Finally, Section 4 offers a conclusion.

2 COMPACT HAND EXTRACTION

Compact hand extraction is divided into four stages: background removal, alignment, finger disassembly and recombination, and illumination compensation.

2.1 Background Removal based on Singular Value Decomposition

The main purpose of background removal is to extract the desired object of an image and remove the unwanted background. The object constitutes the hand that should be kept in an image, whereas the background includes the remaining part of the image, which should be removed. The concepts of object and background are relative, and they depend partly on the specific aims of a research. These aims determine pixel areas that should be modified, partially modified, and unmodified. To extract a hand region for feature extraction, we propose

removing the background by first using singular value decomposition (SVD) (Kalman, 1996)-based image enhancement (SVDIE), and then performing skin color detection (SCD). SVD is a numerical technique for diagonalizing matrices wherein the transformed domain consists of basic states that are optimal to a degree. In general, for any intensity image matrix Ξ_A, $A=\{R,G,B\}$, SVD can be written as

$$\Xi_A = U_A Z_A V_A^T, \quad (1)$$

where U_A and V_A are orthogonal square matrices, and the Z_A matrix represents intensity data and contains the sorted singular values on its main diagonal. Because the hand is located in the foreground of a handshape image and exhibits strong-intensity information, the sub-image of the hand region can be extracted by multiplying Z_A of the original image by an enhancing constant. The ratio of the largest singular value of the generated normalized matrix over an input image is calculated by

$$\xi = \tau \times \frac{\max\left(Z_{gau(\mu=0.5,\sigma=1)}\right)}{\max\left(Z_A\right)}, \quad (2)$$

where $Z_{gau(\mu=0.5,\sigma=1)}$ is the singular value matrix of the synthetic intensity matrix corresponding to the background (with no illumination problems and having a Gaussian PDF with a mean of 0.5 corresponding to the gray level with a value of 127 and a variance of 1 corresponding to the gray level with a value of 32), and τ is a weighting constant that was set at 6 in this study, but may vary across data sets. The obtained ratio ξ was used to regenerate a new singular value matrix, which is actually an equalized intensity matrix of the image generated by

$$\Xi_{e_A} = U_A(\xi \times Z_A)V_A^T, \quad (3)$$

where Ξ_{e_A} represents the equalized image in A color channels. Each image of eight classes is shown in Fig. 1(a), and the ground truth of hand extraction is shown in Fig. 1(b). The result of SVDIE is shown in Fig. 1(c).

In the next stage, we subtracted the SVDIE image from the original one, and the results are shown in Fig. 1(d). To remove the remaining background around the hand, we performed a chrominance-based SCD segmentation with the YC_bC_r color model. The Y value represented the luminance component, whereas the C_b and C_r values represented the chrominance component of the image. When a background-subtracted image, as shown in Fig. 1(d), was presented to the system, a modified YC_bC_r model (Kumar et al., 2003) was applied to build an adaptable skin color cue to enable robust hand detection (Fig. 1(e)). Compared to the manual ground truth shown in Fig. 1(b), our method achieved a high detection rate with a low false alarm, producing a recall rate of 96.46% and a precision rate of 92.51% for 100 images (Table 1).

Figure 1(f) shows that the major drawback of color-based localization techniques is the variability of the skin color footprint under varying lighting conditions, especially for boundary pixels neighboring a dark background. This frequently results in undetected skin regions or falsely detected non-skin textures. As shown in Fig. 1(e), this problem can be resolved using the proposed SVDIE and the SCD method. Figure 1(g) shows the performance of the proposed SVDIE method after residual purging in comparison with Fig. 1(h). For further study, as shown in Fig. 1(i), more holes with different sizes and jagged boundaries were present in the extracted hands of Yin and Xie's work (2007) than in ours. The advantage of the comparison is that it renders the proposed method more suitable for hand extraction for handshape recognition.

2.2 Alignment and Finger Recombination

The need to align, or register, the two hand images is one of the most important steps toward compact hand extraction. This stage involves identifying a spatial mapping that places elements in one hand image into meaningful correspondence with elements in a second hand image. This process is often guided by similarity measures between images that are computed from the image data. However, in time-critical applications, the whole-image-data method for computing similarity is too slow. Instead of using all the image data to compute similarity, a subset of pixels can be used to enhance speed; however, this method may reduce accuracy.

The centroid of an area is similar to the center of gravity of a body. Calculating the centroid involves only the geometrical shape of the area. The center of gravity is equal to the centroid if the body is homogenous (i.e., if it has a constant density). Based on the geometric centroid, the coordinates of the

global centroid (x_G , y_G) can be obtained. Similarly, the coordinates (x_L , y_L) of the local centroid can be calculated. We define the pan or tilt θ of Eq. (4) below as "the angle between the line passing through the global centroid of the whole hand region and the local centroid of the subregion underneath the global centroid."

$$\theta = \tan^{-1}\left(\frac{y_G - y_L}{x_G - x_L}\right) \times 180/\pi \quad (4)$$

The hand image, with rotation in plane θ, can be rotated back to the upright position (orientation alignment). The detailed hand-alignment algorithm, where the region of interest (ROI) is determined to contain as much information as possible (Fig. 2), is summarized as follows:

1). Given a hand detection image (Fig. 2(a)):

2). Calculate the coordinates of the global centroid (x_G , y_G) of the skin region;

3). Block out the ROI with size 100 × 100 pixels, based on the obtained global centroid (Fig. 2(b));

4). Block out the sub-ROI with size 30 × 80 pixels underneath the global centroid (Fig. 2(c));

5). Calculate the coordinates of the local centroid (x_L , y_L) of the segmented subregion (Fig. 2(d));

6). Align the image (Fig. 2(e)), following the calculation of pan/tilt θ (Fig. 2(f)).

2.3 Illumination compensation

Recent research has shown the utility of color in skin detection. To reduce the effect of illumination on color, we further applied SVD to compensate for lighting. To enable partial compensation for variations in lighting without altering the raw data and losing skin color information from the facial image, we suggest adjusting the compensation value dynamically, according to the ratio of the average individual RGB values. This method is highly effective in preserving the skin color data contained in raw images.

Our observations indicated that an overall weighting by fixed-ratio illumination compensation may be unsuitable for all three color channels (R, G, and B). This method restricts the color variation in hand images to a constrained dynamic range, thus failing to display differences between images of different hands and severely affecting recognition. Thus, we devised a method that uses a sliding adjustment of the compensation weightings for each R, G, and B color channel. In this method, the pre-compensation mean for each color channel in the RGB image is first calculated. Using the highest of these channels as a reference value, individual compensation weighting coefficients for the remaining two channels are derived adaptively, according to their ratio to the highest mean. This technique is shown in Eqs. (5)-(8), as follows:

$$\mathrm{Max}(\mu_R, \mu_G, \mu_B) = \eta, \quad (5)$$

$$\lambda_{\mathrm{R}} = \left(\frac{\eta}{\mu_R}\right) \times \frac{\mathrm{Max}\left(Z_{gau(\mu=0.5,\sigma=1)}\right)}{\mathrm{Max}\left(Z_R\right)}, \quad (6)$$

$$\lambda_{\mathrm{G}} = \left(\frac{\eta}{\mu_G}\right) \times \frac{\mathrm{Max}\left(Z_{gau(\mu=0.5,\sigma=1)}\right)}{\mathrm{Max}\left(Z_G\right)}, \quad (7)$$

$$\lambda_{\mathrm{B}} = \left(\frac{\eta}{\mu_B}\right) \times \frac{\mathrm{Max}\left(Z_{gau(\mu=0.5,\sigma=1)}\right)}{\mathrm{Max}\left(Z_B\right)}, \quad (8)$$

where λ_{R} , λ_{G} , and λ_{B} are the individual compensation coefficients for the R, G, and B channels, and μ_{R}, μ_{G}, and μ_{B} are the color means for each channel, respectively. $Z_{G(\mu=0.5,\sigma=1)}$ represents the synthesized normalized intensity image (with no illumination problem and having a Gaussian probability distribution function with a mean of 0.5 and a variance of 1).

3 EXPERIMENTAL RESULTS

In practical applications, significant variations occur among fingers for each class of handshapes. Therefore, the fingers should be disassembled and recombined to remove gaps before proceeding with handshape recognition. We selected six handshape images randomly from our database to examine the proposed methodology. Specifically, as shown in Figs. 3(a)-3(c), the fingers are horizontally and vertically scanned pixel-by-pixel and saved as a more compact form for robust discrimination. The compensation results are shown from left to right in Fig. 3(d). The left column shows the original images, and the middle column shows the images after illumination compensation, using the overall weighting method. Skin color data for same-class hands were concentrated, but data for different-class hands were also extremely concentrated, highlighting overlap among classes being highly deleterious to handshape recognition. The final right column shows images after illumination compensation, using the proposed method, which produced the following results: (1) the effects of variations in lighting were greatly ameliorated; and

(2) skin color data for same-class hands were both more concentrated and similar to those of the original images, whereas skin color data for different-class hands showed a marked difference. These attributes are highly advantageous for handshape recognition.

To analyze the clustering performance of original hands and compensated hands, we used the three leading eigenhands derived from principal component analysis (PCA) to examine their capability to collect similar objects into groups. With three samples per subject, corresponding to Fig. 3(d), Fig. 3(e) shows that the results from compensated hand images were more enhanced than those of hand images without compensation. The proposed method substantially outperformed the overall weighting method in clustering. Figure 3 shows that the method reduced the undesired effects of lighting variances.

4 CONCLUSIONS

A compact hand extraction algorithm for handshape recognition of handshapes has been proposed and tested using our database and video sequences. Based on our SVDIE criteria, this approach performed optimally compared to existing methods. The effectiveness was a result of the ability of the proposed method to recombine fingers and extract hand regions precisely.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support received from NSC through project number NSC 102-2221-E-151-038.

REFERENCES

Farouk, M., Sutherland, A., and Shoukry, 2009. A. A., 2009. A Multistage hierarchical algorithm for hand shape recognition. In *IMVIP 2009 - 13th International Machine Vision and Image Processing Conference*, pp. 106-110.

Thangali, A., Nash, J. P., Sclaroff, S., and Carol, N. 2011. Exploiting phonological constraints for handshape inference in ASL video. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 521-528.

Holt, G. A. T., Reinders, M. J. T., Hendriks, E. A., Ridder, H. D., and Doorn, A. J. V., 2009. Influence of handshape information on automatic sign language recognition. In *Proc. Gesture Workshop*, pp. 301-312.

Murthy, G. R. S., and Jadon, R. S., 2009. *A review of vision based hand gestures recognition. International* Journal of Information Technology and Knowledge Management, vol. 2, pp. 405-410.

Butalia, A., Shah, D., and Dharaskar, R. V., 2010. *Gesture Recognition System. International* Journal of Computer Applications, vol. 1, pp. 61-67.

Khan, I. R., Miyamoto, H., and Morie, T., 2008. Face and arm-posture recognition for secure human-machine interaction. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, pp. 411-417.

Rehrl, T., Bannat, A., Gast, J., Wallhoff, F., Rigoll, G., Mayer, C., Riaz, Z., Radig, B., Sosnowski, S., and K¨uhnlenz, K., 2010. Multiple parallel vision-based recognition in a real-time framework for human-robot-interaction scenarios. In *Proceedings of Third International Conference on Advances in Computer-Human Interactions*, pp. 50-55.

Kim, C., You, B. -J., Jeong, M. -H, and Kim, H., 2008. *Color segmentation robust to brightness variations by using B-spline curve modeling*. Pattern Recognition, vol. 41, pp. 22-37.

Yin, X. and Xie, M., 2007. *Finger identification and hand posture recognition for human-robot interaction*. Image and Vision Computing, vol. 25, pp. 1291-1300.

Jimenez-Hernandez, H., 2010. *Background subtraction approach based on independent component analysis*. Sensors, vol. 10, pp. 6092-6114.

Adan, M., Adan, A., Vazquez, A. S., and Torres, R., 2008. *Biometric verification/identification based on hands natural layout*. Image and Vision Computing, vol. 26, pp. 451-465.

Kakumanu, P., Makrogiannis, S., and Bourbakis, N. G., 2007. *A survey of skin-color modeling and detection methods*. Pattern Recognition, vol. 40, pp. 1106-1122.

Kumar, A., Wong, D. C. M., Shen, H. C., and Jain, A. K., 2003. Personal verification using palmprint and hand geometry biometric. In *Proceedings of the 4th International Conference on Audio- and Video-based Biometric Person Authentication*, pp. 668-678.

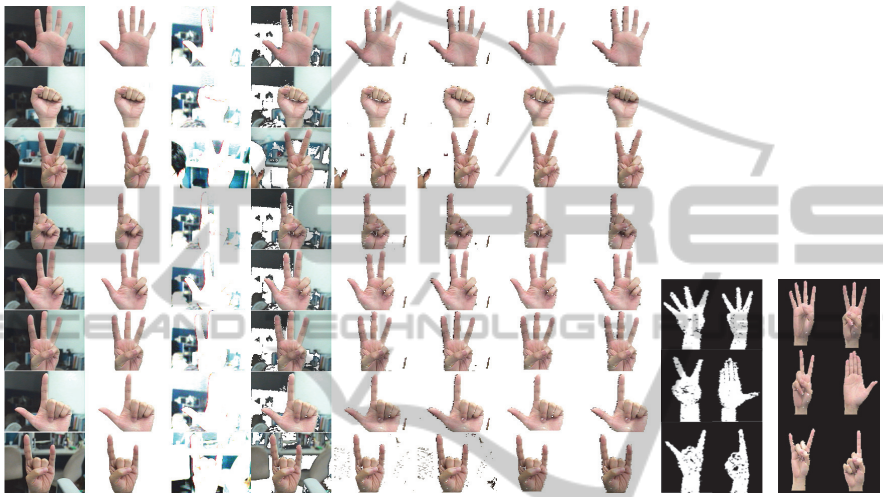
Bakina, I. G., 2011. Person Recognition by hand shape based on skeleton of hand image. Pattern Recognition and Image Analysis, vol. 21, pp. 694-704.

Kalman, D., 1996. *A singularly valuable decomposition: the SVD of a matrix*. The College Mathematics Journal, vol. 27, pp. 2-23.

APPENDIX

Table 1: Performance evaluation by using the recall rate and precision rate with mean and standard deviation.

Statistics \ Rate	Recall (%)	Precision (%)
Mean	96.46	92.51
Standard deviation	0.5514	0.6732

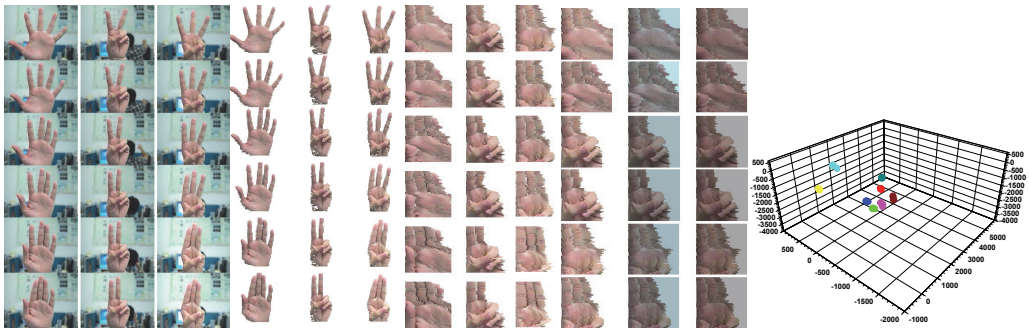


(a) (b) (c) (d) (e) (f) (g) (h) (i)

Figure 1: Background removal with and without the proposed SVDIE method: (a) Input images; (b) Ground-truths; (c) SVDIE images; (d) Background subtraction; (e) SCD with SVDIE; (f) SCD without SVDIE; (g) Residual purging of (e); (h) Residual purging of (f); (i) In comparison with our result (right) to the related work of Yin and Xie's result (2007) (left).

(a) (b) (c) (d) (e) (f)

Figure: 2 Image hand extraction and alignment: (a) Detected hand images, (b) ROI images based on the global centroid, (c) Sub-region underneath the global centroid, (d) Global centroid in red color point and local centroid in blue point, respectively, (e) Variation among handshapes, (f) Aligned images of (d).



(a) (b) (c) (d) (e)

Figure 3: Compact handshape images with lighting compensation: (a) Example images, (b) Aligned handshape images, (c) Compact handshape images, (d) Images of (c) undergone lighting compensation from left to right: original images, overall weighting method, and proposed method, (e) Clustering distribution observed from various view angles corresponding to Fig. 3(d).