Effectual Sampling Scheme for Gesture Recognition

*Jing-Wen Wang

Abstract

This study proposes compact hand extraction to assist in computerized handshape recognition. First, we applied lighting compensation to the adaptable singular value decomposition. then, a tuned sampling algorithm was used to reduce the impact of variations in handshape. We also constructed an auto-eigenhand recognizer with genetic algorithms (GA) for selecting discriminant eigenvector subsets for classification. Although our approach maximized the differences in hand images for various handshapes, it also minimized variations in lighting and pose for the same handshape. Experimental results showed that our method functioned more efficiently than conventional ones that do not use compact hand extraction against complex scenes.

Keywords: compact hand extraction, lighting compensation, tuned sampling

*CorrespondingAuthor: Jing-WeinWang
(E-mail:jwwang@kuas.edu.tw)
Institute of Photonics and Communications, National Kaohsiung University of Applied Sciences, 415, Chien Kung Rd., Kaohsiung, 807, Taiwan

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 close. Handshape recognition in an image poses a challenge because such a 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 [1]-[3]. However, 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 proposed complementary techniques that are 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˚.

2. Sampling Scheme for Recognition

This section presents an approach to handshape analysis, which assumes that each input image contains only one hand and has the potential to recognize eight classes. To verify the performance of the proposed hand posture recognition method, we constructed an integrated system for scaling, translating, and for the rotation-invariant detection of complex scenes. This integration comprises the proposed auto-eigenhand feature, tuned sampling (TS) for feature reduction, and then, a genetically selected auto-eigenhand recognizer for obtaining the most discriminative feature subset accompanied by the acceptance/rejection threshold value. The independent principal component analysis (PCA), also named the auto-eigenhand technique, was performed for each subject using the available TS images. The result of this analysis is a set of eigenhands for each subject. We described the selected auto-eigenhand subset that maximizd the distances between the hand images of different handshapes, and minimized the distance between the hand images of the same handshape.

2.1 Tuned vs. log-polar sampling

Variations in hand pose are major factors in the feature distribution of handshape recognition. These variations result in excessively scattered spatial distributions for hand images of the same class, and overlap in spatial distributions for hand images of different classes. An objective of this study was to obtain more concentrated feature-space distribution for the same hand image class, and simultaneously widen feature-space distribution for dissimilar hand image classes. Recent researches [4]-[5] used log-polar sampling (LPS) to overcome the problem of head pose variation and to obtain scale-invariant, translation-invariant, and rotation-invariant results. Using LPS on an image involves sampling with the space-varying log-polar grid and constructing a corresponding log-polar image. The log-polar grid consists of N_r concentric circles with radii increasing exponentially from the center to the periphery, and N_θ uniformly spaced angular sampling points. The sampling density increases from the periphery to the center grid. A sampling point (r_i,θ_i) on the log-polar image maps back to $(x_i = e^{r_i}\cos\theta_i, y_i = e^{r_i}\sin\theta_i)$ on the Cartesian coordinate plane. At each of these sampling points (x_i,y_i), a sampling function is defined to cover a circular patch on the original image with the diameter given by

$$D = \frac{2\pi\sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}}{N_\theta} \quad (1)$$

where (x_c, y_c) is the center of the sampling. The intensity value at (r_i, θ_i) is considered the mean pixel value within the circular patch. The sample spacing in the log-polar image is given by

$$\delta_r = \frac{\log R_{\max} - \log R_{\min}}{N_r} \tag{2}$$

$$\delta_\theta = \frac{2\pi}{N_\theta} \tag{3}$$

where $R_{\max}$ and $R_{\min}$ are the radii of the largest and smallest sampling circles respectively. The log-polar feature vector is formed by either column or row concatenation of the log-polar image. The radii of LPS circles increase exponentially from the center to the periphery. Even a small variation introduced by the distance from the center point would cause nonlinear amplification, and the resulting sampling image would be substantially altered.

Because of this weakness in the model, we proposed a novel tuned sampling (TS) algorithm as a solution. The TS algorithm is beneficial because it uses (1) the Cartesian coordinate system, thereby avoiding the nonlinear amplification of changes over small distances associated with the log-polar sampling system, and (2) a tuned sampling technique with alternating square and diamond patterns, which can be effective in compensating for the problem of variance in the sampling mean produced by translation in hand images. At each of these sampling points, the value is represented by averaging the window pixels with diameter *D* surrounding the tuned point. The TS image comprises*N* tunes of various sizes. The tuned sampling density is given by

$$r = \frac{Min(T_w, T_h)}{N} \tag{4}$$

where r is the distance of sampling points; T_w and T_h are the width and height of handshape images, respectively; N is the number of total tunes, where N is set to 25 regarding the hand-image size. The TS image obtained from the handshape image function $\zeta(\cdot)$ is given by

$$m = \frac{n}{c} + 1 \tag{5}$$

$0 \le k \le 7,\ 0 \le n < 50$

$$t(n,k) = \frac{1}{(2m+1)^2}\left(\sum_{w=-m}^{m} \sum_{h=-m}^{m} \zeta \begin{pmatrix} (n+1) \times r \sin(k \times \frac{\pi}{2}) + w \\ (n+1) \times r \cos(k \times \frac{\pi}{2}) + h \end{pmatrix} \right) \tag{6}$$

where $t(n,k)$ is the TS image; n is the number of tunes; k is the number of the tuned points; m is an integer; and c is the sampling density, where c is set to 10 so that the sampling is not too dense, thus resulting in feature overlapping.

2.2 Auto-eigenhand selection

In this study, genetic algorithms (GAs) [6] were used to evaluate the effectiveness of an auto-eigenhand subset and to search for the best *d* components from the given auto-eigenhandset with *n* eigenvectors. The basic concept of a GA is to maintain a population of knowledge structures (chromosomes) that represents a pool of possible solutions for the problem of optimization. In this application, each structure of the population represents a possible set of parameters for the system to be optimized. The GA then searches for thestructure that is the best fit in the optimum setting of parameters. A straightforward approach for

representing the subsets of a given set with (n + 3) elements in a binary chromosome **c** = $\langle c_1, c_2, ..., c_n, \xi_1, \xi_2, \xi_3 \rangle$ is to index the set's elements and associate a position in the chromosome with the element's index. Here, the combination of ξ_1, ξ_2,and ξ_3 represents the setting of the threshold in the fitness function. Therefore, if a feature that corresponds to the description of an eigenvector on a hand property is included in the selected subset, component c_i of the corresponding chromosome **c** is 1; otherwise, c_i = 0.

The fitness of chromosome **c** can be defined as the classification rate of its corresponding feature set. First of all, we randomly generated an initial population. After evaluating the fitness of each chromosome in the population, a selection procedure determines the chromosomes that are part of the reproduction process. As in natural evolution, survival of the fittest is the main strategy. Next, the selected chromosomes are altered by using a mutation operator to form new fit chromosomes from fit parents. A mutation operator with probabilities that vary according to the step function serves as a secondary search operator, ensuring the reach ability of all points in the search space. Random changes of selected chromosomes generate candidates for the solution to the optimization problem in various regions of the search space. The resulting offspring, with the newborn chromosomes, maintains the population size and is then evaluated and inserted back into the population. This process continues until either an absolute fittest chromosome is detected in the population or a predetermined maximum number of generations is reached.

As outlined schematically with the leave-one-out learning algorithm [7], all but one of the image sets were used as the training set, and the one excluded was used for testing the performance of the Mahalanobis classifier. For inside testing, each input image is asked to reveal its serial number for performance evaluation.In contrast, no concomitant identity information is required for outside testing. Fitness function is critical to the performance of a GA. In our GA approach, we adopted a Bayesian likelihood function as the fitness function to explore the importance of individual features in the optimal classification. The formula is described as

$$1 = (1 - d/n) \times \frac{\rho_1 \times AAR - \rho_2 \times FRR}{|\rho_3 \times ARR - \rho_4 \times FAR|} \quad (7)$$

The ratio d/n of the selected feature number to the total feature number is known a priority. Four possible outcomes exist in a recognition system operating in recognition mode: accurate acceptance rate (*AAR*), accurate rejection rate(*ARR*), false rejection rate (*FRR*), and false acceptance rate (*FAR*). An *FAR* occurs when the system incorrectly matches a sample to a template, thereby incorrectly identifying a handshape and potentially allowing unauthorized individuals (or impostors) to gain access. An *FRR* occurs when the biometric system incorrectly rejects a valid sample, thus denying access to a legitimate hand. Setting rejection thresholds at high levels enhances security, but may affect usability because of a high *FRR*. Conversely, low acceptability thresholds may compromise security because of a high *FAR*. The ratio of the number of selected features to the total number of features is also known a priority. The evaluation rule

of the likelihood ratio, derived from Bayes' decision rule for the minimum error classification rate, was devised to explore the importance of individual features in optimal recognition, and can be interpreted as maximizing the *AAR* and *ARR* while minimizing the *FRR* and *FAR*. The novelty of the presented recognizer is based on its particular merit,which is the derivation of the novel likelihood fitness function for the selection of discriminative features with a concomitant threshold for recognition. Using the Bayesian likelihood fitness function, these outcomes are made suitable for any desired working point by altering the action factors (ρ_1 , ρ_2 , ρ_3 , and ρ_4). In this study, the action factors were determined experimentally and set to $\rho_1 = 10$, $\rho_2 = 1$, $\rho_3 = 1$, and $\rho_4 = 10$. The selection algorithm is as follows:

1).Perform TS and LPS on the segmented hands, and then array these subimages in a matrix $\boldsymbol{x} \in R^{\zeta}$ to obtain auto-eigenhands, where ζ is the correspondingimage size, with one column per sample image.

2). Select ϖauto-eigenhand components and generate the projection vector for each of the training subject vectors $\boldsymbol{x}_{jl}$.

$$\boldsymbol{y}_{jl} = \boldsymbol{W}_{\varpi}^{k\,T} \boldsymbol{x}_{jl} \tag{8}$$

where $\boldsymbol{x}_{jl}$, j = 1, 2,…, J and l = 1, 2,…, L, is the lth image of the jth subject, $\boldsymbol{W}_{\varpi}^{j}$ = ($\boldsymbol{w}_1^j$,…, $\boldsymbol{w}_{\varpi}^j$), and $\varpi < n$. The vector $\boldsymbol{w}_{\varpi}^j$ is the auto-eigenhand corresponding to the ϖth auto-eigenhand selected from the sample covariance matrix of subject j.

3). Calculate the cosine Mahalanobis distance function for subjects in the database using the formulas

$$\boldsymbol{y}_z = \boldsymbol{W}_{\varpi}^{j\,T} \boldsymbol{x}_z \text{ , } 1 \le z \le \varpi \tag{9}$$

$$\varsigma_{jz} = \min_{z} \frac{\left\| y_z - m_j \right\|^2}{\left\| v_j \right\|} \tag{10}$$

$$\gamma = arg(\min_{j}(\sum_{z=1}^{Z} \varsigma_{jz})) \text{ and } \varsigma_{jz} < \delta \tag{11}$$

where $\boldsymbol{y}_z$ is the projection vector of the zth test auto-eigenhand vector, and ς_{jz} is the distance of the test hand image z from the j th category. The mean $\boldsymbol{m}_j$ and the variance $\boldsymbol{v}_j$ of the auto-eigenhand from the jth category are calculated with the leave-one-out cross validation. The category label is denoted as γ, and δ stands for the recognition threshold. For inside testing, there are three types of recognition results: *AAR*, *FAR*, and *FRR*. The *ARR* and *FAR* are measured for outside testing.

3. Experiment and Analysis

We performed a set of compact hand extraction and handshape recognition experiments to demonstrate the efficiency of the proposed method. We used our laboratory database for testing because no public database has been available to date. The database contains 800 single light source images of 100 hands, each seen under eight handshapes. For each hand extraction that represents a single posture, a SonyHAD Color Video Camera was set up to capture static images across various handshapes under office luminance. From 12 subjects in our

laboratory database, we randomly selected 96 images including eight-class handshape images from each person. Of these eight-class images, are shown in Fig. 1(a).We randomly selected seven-class handshape images and grouped them into the database; the remaining one-class handshape images were assigned with an “invader” status. For each of our selected 11 database members, we prepared eight-class images, including eight original images and eight synthetic images from each original, producing 704 images. For the invader, we prepared eight images for each class, comprising 64 invader images. The synthetic images were regarded as supplementary, generated to achieve greater generalization capability, as shown in Figs. 1(b) and 1(c). They were generated by rotating each raw image -30°, -15°, -10°, 10°, 15°, and 30°respectively, resulting in six rotated images, and by adding 3% and 5% Gaussian noise to the original image, totaling eight synthetic images for each raw hand image. The dimensions of all training and testing images were 160 × 120 pixels.

(i) (ii)

(a)

(b)

(c)

Figure1: (a) Our 8-class handshape images: (i) 7-class database images; (ii) Invader images, (b) Gaussian noise images, (c) Rotated images.

Handshape recognition testing was performed by using the proposed framework. For each input image, compact hand extraction and lighting compensation were first conducted. The tuned image and log-polar sampling image were then subjected to eigenhand and auto-eigenhand analysis for feature extraction. Finally, using GAs with leave-one-out validation under fitness guidance and after extensive recursive evolution and recombination, optimal feature sets for use in recognition were derived. Parameters for the designed GA were determined experimentally: population size = 20, number of generations = 1000, and the probability of crossover = 0.5. The mutation probability value started with a value of 0.1 and was varied as a step function of the number of iterations until it reached a value of 0.001.The objectives of this stage were twofold. First, we compared the performance of the auto-eigenhand technique with and without compact hand extraction. Second, we compared the recognition accuracy of the LPS and TS algorithm by running several experiments using inside and outside image samples, respectively, and singling out auto-eigenhand components using GAs. The GA procedure accompanying the initial threshold is repeated until only the most consistent auto-eigenhand subset, with a corresponding threshold, is acquired. The resulting average of three independent runs for both LPS and TS was reported. We compared the results obtained by LPS recognition and the corresponding results for TS. The results, shown in Table 1, indicated that because the experiments included the testing of database and live images, the recognition achieved by TS declined at a lower rate than that achieved by LPS. In the experiments, LPS achieved a slightly higher AAR rate than did TS. This result was due to the discrepancy between the two image resolutions, with LPS having a resolution of 40×35 pixels and TS a low resolution of 50×8 pixels.

Table 1.Handshape recognition of LPS vs. TS with compact hand extraction: Eigenhand vs. auto-eigenhand features.

Recognition Rate (%) / Method	AAR	ARR	FAR	FRR
LPS + eigenhand	78.46	96.33	0.0255	0.217
LPS + auto-eigenhand	96.51	97.82	0.014	0.0039
TS + eigenhand	85.20	97.48	0.0108	0.168
TS + auto-eigenhand	97.67	98.33	0.008	0.0026

4. Conclusion

To evaluate the feasibility of recognizing the class of a handshape in a complex environment, the system was tested on our database of 704 images and 64 intruders for frontal and near-frontal views. Simulation results showed that the proposed framework produced excellent results for recognition accuracy. Additionally, experimental results for the live images demonstrated the effectiveness of the proposed framework and confirmed that our system was reliable for real-time recognition. Based on these results, our framework can be applied to gesture modeling and recognition systems [8] in two ways: each handshape can be processed separately or through tracking information.

Acknowledgment

This study is supported in part by the National Science Council of the Republic of China under contract number NSC 101-2221-E-151-069.

References

[1] M. Farouk, A. Sutherland, and A. A., Shoukry, “A multistage hierarchical algorithm for hand shape recognition,” IMVIP 2009 - 13th International Machine Vision and Image Processing Conference, pp. 106-110, 2009.

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

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

[4] L. H. Koh, S. Ranganath, and Y. V. Venkatesh, “An integrated automatic face detection and recognition system,” Pattern Recognition, vol. 35, pp. 1259-1273, 2002.

[5] S. Arivazhagan, K. Gowri, and L. Ganesan, “Rotation and scale-invariant texture classification using log-polar and ridgelet transform,” Journal of Pattern Recognition Research, vol. 1, pp. 131-139, 2010.

[6] M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn, and A. K. Jain, “Dimensionality reduction using genetic algorithms,” IEEE Transactions on Evolutionary Computation, vol.. 4, pp. 164-171, 2000.

[7] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, 2nd ed. Wiley-Interscience, 2000.

[8] M. M. Hasan and P. K. Mishra, “Hand gesture modeling and recognition using geometric features: a review,” Canadian Journal on Image Processing and Computer Vision, vol. 3, pp. 12-26, 2012.

Jing-Wein Wang received the B.S. and M.S. degrees in Electronic Engineering from National Taiwan University ofScience andTechnology, in 1986 and 1988, respectively, and the Ph.D. degree in Electrical Engineering from National Cheng Kung University, Taiwan, in 1998. From 1992 to 2000, he was a principal project leader at Equipment Design Center of PHILIPS, Taiwan. In 2000, he joined the faculty of National Kaohsiung University of Applied Sciences, where he is currently a professor and dean in the Institute of Photonics and Communications. His current research interests are combinatorial optimization, pattern recognition, wavelets and their applications.