Scale Invariant static hand-postures detection using Extended Higher-order Local Autocorrelation features

Isack Bulugu
Department of Electronic Engineering
and Information Science,
University of science and
Technology of China

Zhongfu Ye
Department of Electronic Engineering
and Information Science,
University of science and
Technology of China

ABSTRACT
This paper presents scale invariant static hand postures detection methods using extended HLAC features extractedfrom Log-Polar images. Scale changes of a handposture in an image are represented as shift in Log-Polar image. Robustness of the method is achieved through extracting spectral features from theeach row of the Log-Polar image. Linear Discriminant Analysis was used to combine features with simple classification methods in order to realize scale invariant hand postures detection and classification.The method was successful tested by performing experiment using NSU hand posture dataset images which consists 10 classes of postures, 24 samples of images per class, which are captured by the position and size of the hand within the image frame. The results showed that the detection rate using Extended-HLAC can averaged reach 94.63% higher than using HLAC features on a Intel Core i5-4590 CPU running at 3.3 GHz.

General Terms
Scale invariant, HLAC features, log polar image, hand posture, linear discriminant analysis, posture detection, posture classification.

Keywords
Scale invariant, log polar image, posture detection, posture classification.

1. INTRODUCTION
Hand-gesture-based methods stand out from other approaches by providing a natural way of interaction and communication [1]. Ong and Ranganath [2] presented a thorough review on hand gesture analysis in relation to the problem associated with it. The hand gesture recognition are based on the hand shape (*static gesture*) or the movement of the hand (*dynamic gesture*).Various methods using basic features such as gray-level histograms, Fourier features and Gabor features have been proposed in the last few decades as an improvement to the static hand-gestures recognition analysis. Among them, we specifically address extended higher order local autocorrelation (HLAC) features [11] to hand gestures recognition. Higher-Order Local Autocorrelation features [3] have got many advantages for their wide availability to image analysis. They have performed very well in face recognition [4] and [5], natural object recognition [6], and gesture recognition [7].

In order to avoid the effect of changes in position of hands, hand-gesture recognition is divided into two stages; detection and classification. Usually hand's position is searched in detection stage and then the hand centered at that position is easily classified into classification stage.

Robustness of hand detection is very crucial toward changes of scale since the size of hands are not similar in all the images. This methods was clearly described in the face detection methods [8] [9] [10]the scale changes of the face is coped with by changing the size of the image itself. Therefore this work use the same method successful in hands detection.

2. SCALE INVARIANT FEATURES
Based on the simplest models of space variant sensor that is Log-polar transformation [13], the input image is given higher weight at the central region rather than the peripheral region. Therefore, log-polar image is the best for target recognition since peripheral region contains mostly background information. Furthermore, scale changes and rotations of a target are represented as shifts in the log-polar image if the center of the target is fixed at the center of the image. It implies that we can get scale and rotation invariant features by extracting shift invariant features from log-polar images. Prior to this work, based on these properties [13] they proposed scale and rotation invariant features which are based on Higher-order Local AutoCorrelation (HLAC).Therefore, we used the same properties to proposed scale and rotation invariant based on Extended Higher-order Local AutoCorrelation.

2.1 Log polar transformation
Log-polar image can be constructed by transformation of Cartesian coordinates to polar coordinates. Input image is generally represented as a collection of pixel points on the Cartesian coordinate. Consider point (x,y)on the cartesian coordinate is transformed into point on the polar coordinate. Then the point in polar coordinate is transformed into the point (z=log (ρ), θ) on the log-polar coordinate.

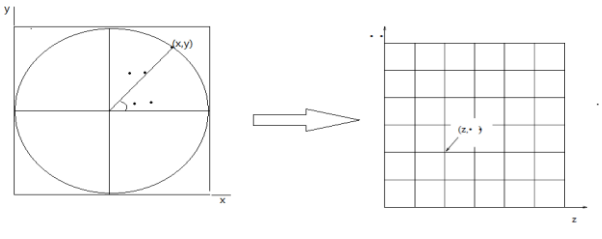

Figure1. Transformation from Cartesian coordinate to Log-polar coordinate.

2.2 Extended HLAC features
The N th-order autocorrelation functions, extensions of autocorrelation functions, are defined as $x(a_1,a_2,\ldots\ldots.a_N)=(\int f(r)+f(r+a_1)\ldots\ldots..f(r+a_N))$

Where $f(r)$ denotes the intensity at the observing pixel r, and$a_1,a_2,\ldots\ldots.a_N$are$N$ displacements.

The extension of HLAC features was presented by T. Tayoda and O.Hasegawa [11]. They extend the original HLAC features which were restricted up to the second order and are extracted by 25 mask patterns. Advantages of Extended-HLAC features over others are such as they increase orders up to eight and extract the HLAC features using 223 mask patterns. Furthermore, they create large mask patterns and construct multi-resolution features to support large displacement regions. Therefore, images are more closely characterized using various mask patterns.

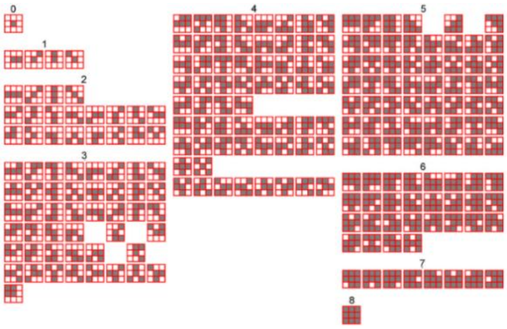

Figure 2. 223 mask pattern of the zeroth-order to eighth-order HLAC features (3 × 3 pixels)

The Extended HLAC feature values of an image are calculated by scanning the image with the 223 mask patterns in Fig. 1 and computing the sums of products of the intensities of corresponding pixels. Each value represents the power spectrum of the mask pattern. Therefore, the 223 mask patterns are regarded as the basis functions of frequency analysis.

It is well known that Extended HLAC features are shift invariant. This means that scale and rotation invariant features can be obtained by extracting shift invariant features from Log-Polar image along horizontal axis. The features of a 3×3 displacement region mainly extract the local detailed information. Then each mask pattern in Fig.1 is scanned over the entire image to calculate HLAC feature value *f*. For each input image, the operation is performed using 223 different mask patterns to create the feature vector (f_1, f_2, …, f_{223}) in which $f_1$denotes the total length of polygon boundary. With the total area of a gesture pattern as another feature f_0, combining with the 223 dimensional HLAC based feature vector (f_1, f_2, …, f_{223}), we create a 223 dimensional feature vector (f_0, f_1, f_2, …, f_{223}) for image contains one gestures.

2.2.1 *Feature extraction*

By using auto-covariance of the input signal *x(t)* is define as

$$R(s) = \frac{1}{N}\sum_{t=0}^{N-1}(x(t) - \bar{x})(x(t+s) - \bar{x})$$

Where $x(t)$ depends on only the difference.The autocorrelation $\rho(s)$ of signal $x(t)$ is defined by

$$-1 \le \rho(s) = \frac{R(s)}{R(o)} \le 1$$

At s=0, autocorrelation is maximum and becomes robust to scaling of the signal because it is normalized by the variance *R(0)*. It implies that features based on the autocorrelation are robust to the scale changes of the intensities.

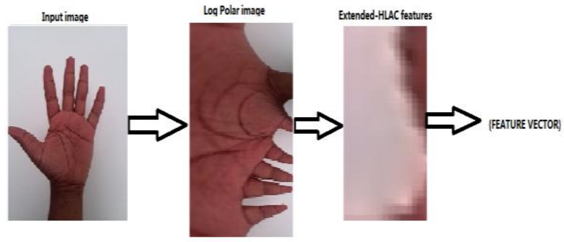

Figure 3. Extended-HLAC Features

3. HAND GESTURE RECOGNITION

The extracted features from a Log-Polar image are general and primitive and are independent on the recognition task. We expect that these features have enough information to distinguish hand gestures. In order to perform recognition task are formed by combining these features by using Linear Discriminant Analysis (LDA). For detection task, we have to design a classifier which can classify "hand" and "not hand". We expect that "hand" class to have only hand images which includes all hand gesture images such as but "not hand" class includes many kinds of images except hand images. It is difficult to recognize "not hand" class as a single cluster in the feature space. Thus, by modifying the discriminant criterion such that the covariance of "hand" class is minimized while the covariance between "hand" class and each of the learning samples in "not hand" class is maximized.

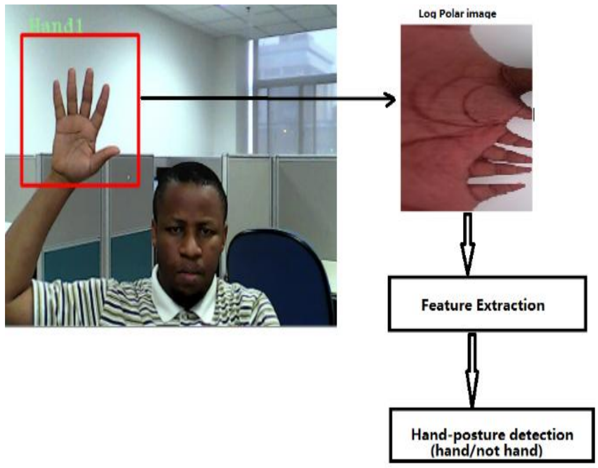

Figure 4. The flow of hand posture detection method

$\boldsymbol{C_H} = \{\boldsymbol{X_{HFi}}/i=1,\ldots\ldots.\boldsymbol{N_H}\}$

$\boldsymbol{C_{NH}} = \{\boldsymbol{X_{NHi}}/i=1,\ldots\ldots.\boldsymbol{N_{NH}}\}$

Where $\boldsymbol{C_H}$is the number of hand image samples and $\boldsymbol{C_{NH}}$ is the number of images with no hands. Then $\boldsymbol{X_H}$ represents mean vectors of Hands class, $\boldsymbol{C_H}$.

$$\boldsymbol{X_H} = \frac{\boldsymbol{1}}{\boldsymbol{N_H}}\sum_{i}^{1}\boldsymbol{X_{Hi}}$$

The covariance matrix (Σ_H) of "Hand" class and the covariance matrix (Σ_C) between the mean vector of "hand" classand each samples of "not hand" class are given by

$$\sum\nolimits_H = \frac{\boldsymbol{1}}{\boldsymbol{N_H}}\boldsymbol{X_{Hi}}\boldsymbol{X_{Hi}}^{\boldsymbol{T}} - \overline{\boldsymbol{X_{Hi}}}\,\overline{\boldsymbol{X_{Hi}}}^{\boldsymbol{T}}$$

$$\sum\nolimits_C = \frac{\boldsymbol{1}}{\boldsymbol{N_{NH}}}\sum_{k}^{\boldsymbol{N_{NH}}}(\boldsymbol{X_{HNi}} - \overline{\boldsymbol{X_{Hi}}})(\boldsymbol{X_{HNi}} - \overline{\boldsymbol{X_{Hi}}})^{\boldsymbol{T}}$$

New features $\boldsymbol{y} = (y_1, \cdot \cdot \cdot, y_L)^T$are obtained by linear combination of primitive features $\boldsymbol{x} = (x_1, \cdot \cdot \cdot, x_M)^T$ as

$\boldsymbol{y} = \boldsymbol{A^T x}$where$A = [aij]$ is a coefficients matrix, L is the number of new features, and M is the number of primitive features. Next step, we use discriminant criterion (J) to construct discriminant space in which the covariance of "face" class is minimized and the covariancebetween the mean vector of "face" class and each samples of "not face" class is maximized.

$$J = \sum\nolimits_H^{-1} \sum\nolimits_C$$

The optimal coefficient *A*, which maximizes this discriminant criterion *J*, is obtained by solving the eigen-value problem

$A^T \Sigma_H A = I$

Learning samples are used to construct discriminant space for 'Hand' and 'not hand' classification. If we have less distance for the mean vector of 'Hand' from a given threshold, then that position is classified as 'hand'. In this classification, hand detection depends much on threshold value. The optimal value of threshold can be determined by using the following two probabilities;

$$(i) \quad P_H = 1 - \frac{n_H}{N_H}$$

Where P_H is a probability in which the samples of "hand" class aremiss-classified as "not hand",$n_H$is the number of the samples of "face" class which has a value less than thethreshold and $N_H$is the total number of samples of "face" class.

$$(ii) \quad P_{NH} = 1 - \frac{n_{NH}}{N_{NH}}$$

$where\ P_{NH}$ Is the probability in which thesamples of "not hand" class are miss-classified as "hand",$n_{NH}$is the number of the sample of "not hand" classwhich has a value less than the threshold and $N_{NH}$is the total number of samples of the "not hand" class.

As thethreshold is increasing from zero to infinity, two probabilities may change depending on the threshold. Since thesetwo probabilities are error probabilities, we would like to minimize both of these probabilities. Thus we can selectthe optimal threshold in which the sum of the two probabilities is minimized.

4. EXPERIMENT

In following experiment, we use NSU hand posture dataset I [14] which consists 10 classes of postures, 24 samples of images per class, which are captured by the position and size of the hand within the image frame. Also, we performed the experiment to evaluate the robustness to the scale changes of the hand postures using NSU hand posture dataset II [15] which include images with only background, Hand postures with other noise such as books, chair, clothes, and other parts of human body. Some of the images from the two dataset with hand postures without background noise and with background noise images and their Log-Polar image are shown in Figure 5.It was discoveredthat Log-Polar image is robust to the scale changes of hand palm and background changes.

We have performed the experiment to evaluate the performance of the hand posture detection method using HLAC features, Fourier power spectrum features and Extended-HLAC features. The 200 images are selected atrandom from the dataset of the class contains hand posture images. The correct position of the "hand" is measured in advance for all ofthe selected images. These selected images are used to evaluate the performance of the posture detection.

Precision of the static hand posture detection, it can be investigated by evaluating the correct rate by changing the permission distance between the detected position and the correct position. Table 2 shows the correct detection rates when the permission distance is changed from 0.0 to 5.0.

Figure5 (a). The examples of hand-gesture images without background noise and their Log-polar images.

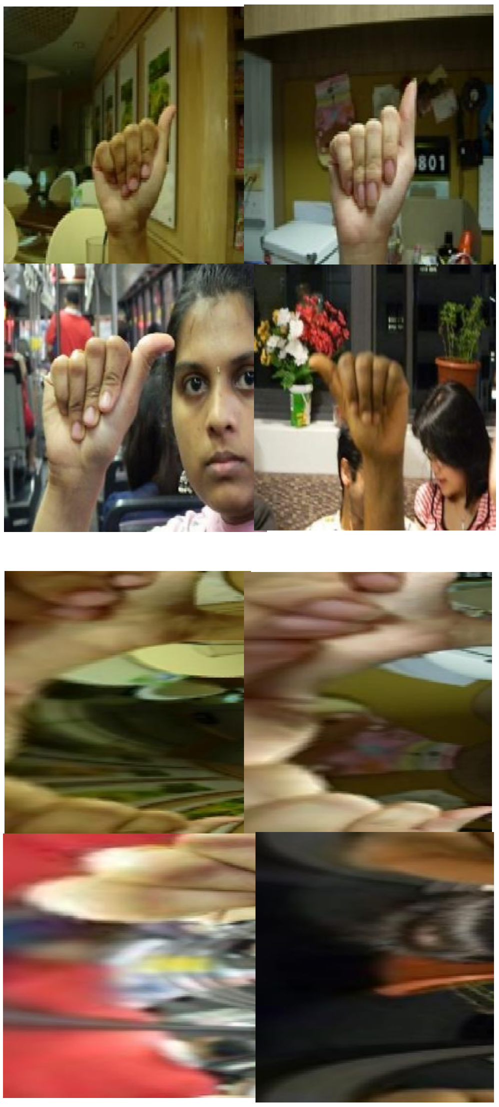

Figure 5(b).The examples of hand-gesture images with background noise and their Log-Polar images.

5. RESULTS AND DISCUSSION

Below are some of results obtained by using a proposed method. It has been tested and showed great performance with images of different background, noise in background and other parts of human body.

Table 1. The performance of hand-posture detection method for 200 images

	Gesture detected
HLAC	180/200
Fourier Power Spectrum	192/200
Extended-HLAC (Our work)	197/200

It was found that the recognition rates of the proposed method is higher than the previous method based on Extended-HLAC features. This means that the proposed method is improved by dropping the 2D rotation invariance. The changes of the resolution axis is less influenced to the recognition rate.

6. CONCLUSION

We proposed scale invariant static hand-postures detection methods using extended higher-order local autocorrelation features extractedfrom Log-Polar images. We increased their orders up to eight and represented the extended HLAC features with 223 mask patterns. By this extension, the image is characterized more closely than conventional second-order features. In hand posture detection, the proposed method demonstrated high detection rate of 98.38% which far better than previous methods.

The proposed method is good for practical applications because its features are computed rapidly. In future, its speed can be enhanced further by implementation of the vision chip.

Table 2. Recognition rate according to different resolution

	32X32	64X64	128X128	256X256
Extended-HLAC(our work)	97.20%	98.31%	98.45%	99.56%
Fourier Power Spectrum	96.1%	97.24%	98.20	98.57%
HLAC	77.88%	79.45%	83.23%	85.21%

7. ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards making this paper more readable, and their insightful comments and suggestions are more appreciated.This work is supported by University of Science and Technology of China and CAS-TWAS Presidents Fellowship.

8. REFERENCES

[1] Vladimir I. Pavlovic, Rajeev Sharma, Thomas S. Huang Visual interpretation of hand gestures for human–computer interaction IEEE Trans. Pattern Anal. Mach. Intell., 19 (6) (1997), pp. 677–695

[2] S.C.W. Ong, S. RanganathAutomatic sign language analysis: a survey and the future beyond lexical meaning IEEE Transactions on Pattern Analysis and Machine Intelligence, 27 (6) (2005), pp. 873–891 View Record in ScopusFull Text via CrossRefCiting articles (223)

[3] N. Otsu, T. Kurita, A new scheme for practical flexible and intelligent vision systems, in: Proceedings of the IAPR Workshop on Computer Vision, 1988, pp. 431–435.

[4] T. Kurita, N. Otsu, T. Sato, A face recognition method using higher order local autocorrelation and multivariate analysis, in: Proceedings of the International Conference on Pattern Recognition, vol. 2, 1992, pp. 213–216.

[5] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma, N. Otsu Face recognition system using local autocorrelations and multiscale integration IEEE Trans. Pattern Anal. Mach. Intell., 18 (10) (1996), pp. 1024–1028

[6] M. Kreutz, B. Völpel, H. Janssen Scale-invariant image recognition based on higher order autocorrelation features Pattern Recognition, 29 (1) (1996), pp. 19–26

[7] T. Kurita, S. Hayamizu, Gesture recognition using HLAC features of PARCOR images and HMM based recognizer, in: Proceedings of the International Conference on Automatic Face and Gesture Recognition, 1998, pp. 422–427.

[8] K.Sung and T.Poggio, "Example-based learning for view-based human face detection," tech. rep., A.I. Memo1521, CBCL Paper 112, 1994.

[9] H.A.Rowley, S.Baluja, and T.Kanade, "Human face detection in visual scenes," tech. rep., CMU-CS-95-158R,1995.

[10] H.A.Rowley, S.Baluja, and T.Kanade, "Rotation invariant neural network-based face detection," tech. rep., CMU-CS-97-201, 1997.

[11] T. Toyoda,O. Hasegawa, "Extension of higher order local autocorrelation features"Pattern RecognitionVolume 40, Issue 5, May 2007, Pages 1466–1473

[12] B.Moghaddam and A.Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. on PatternAnalysis and Machine Intelligence* 19(7), 1997.

[13] L.Massone, G.Sandini, and V.Tagliasco, "Form-invariant: Topological mapping strategy for 2d shape recognition,"*Computer Vision, Graphics and Image Processing* 30, pp. 169–188, 1985.

[14] P.Kumar, P.Vadakkepat, L.Poh, "Hand Posture And Face Recognition using a Fuzzy-rough approach" International Journal of Humanoid Robotics Vol. 7, No. 3 (2010) 331–356

[15] P.K.Pisharady, P.Vadakkepat, A.P.Loh, "Attention based detection and recognition of hand postures against complex backgrounds", Int.J.Comput.Vis.101 (3) (2013)403–419.