Static Hand Gesture Recognition Using Discriminative 2D Zernike Moments

Md Abdul Aowal[†], Adeeb Shahriar Zaman[†]
S M Mahbubur Rahman[*†], *Member*, *IEEE*, Dimitrios Hatzinakos[‡], *Senior Member*, *IEEE*
[†]Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh
[‡]Identity, Privacy and Security Institute, Department of Electrical and Computer Engineering
University of Toronto, Toronto, ON, Canada, M5S 2E4
E-mails: {aowal.eee, adeeb.math}@gmail.com, mahbubur@eee.buet.ac.bd, dimitris@comm.utoronto.ca

***Abstract*—Hand gesture recognition plays a vital role in developing vision-based communication for human-computer interaction. This paper presents a novel static hand gesture recognition method using the two dimensional Zernike moments (2D ZMs) those are considered as effective features when patterns in images possess distortions due to rotation, scaling or viewing angle. The key contribution of this paper lies in the fact that a discriminative set of ZMs are used to represent features of the hand postures as opposed to traditional features obtained from heuristic choice of fixed-order moments. The orthogonal nature of the 2D ZMs allows the estimation of the discrimination power of the individual moments by using the inter- and intra-class variances of the features. The nearest neighbor classifier is employed on the discriminative ZMs (DZMs) to recognize the hand postures in a computationally efficient way. Experimental results on commonly-referred database show that the proposed DZM-based method provides recognition accuracies better than that provided by the conventional principal component analysis, Fourier descriptor or existing ZM-based methods.**

***Keywords*—Discriminative moments, hand gesture recognition, Zernike moments**

I. Introduction

In the area of human computer interaction (HCI), many researches have focused on developing vision-based advanced hand gesture recognition schemes. This is mainly due the fact that unlike traditional HCI devices such as mice, key boards and gloves, hand gestures are less intrusive and more convenient for natural way of communication for humans. The potential applications of hand gesture include the HCI in robotics, assistive systems, sign language communication and virtual reality [1].

In general, hand gesture deals with the motion trajectories of the moving hands with specific shapes defined in standard vocabulary. A very few research studies consider three dimensional (3D) model to recognize hand gesture using tracking parameters [2]. Due to the low computational complexity with appreciable recognition accuracy, communication between computer and humans through static hand gesture has become increasingly popular. In the literature, the 'static hand gesture' is very often referred to as the 'hand posture' that considers 2D models such as the image contour and the silhouette. Hand posture recognition methods face challenges due to the large number of degrees of freedom of inter-personal human hands that can deform extensively and non-ideal imaging environments such as illumination variations, inconsistencies of viewing angles and scale variations due to changeable distance of camera. Due to these variabilities, in practice, the dimensionality of the shapes of the hand postures is reduced to a small number of features that are effective for recognition.

In the literature of hand posture recognition, a number of feature sets such as those based on the principal component analysis (PCA) [1], radial signature [3], centroid distance signature, histogram of gradients, local binary patterns, Shi-Tomasi key corners [4], scale invariant feature transform, speeded up robust features (SURF), Fourier descriptors (FD) [5], invariants of geometric moments [6], and orthogonal moments [7], [8] have been employed to represent the shapes or silhouettes of hand postures. In [9], the Hu moments and SURF features are combined, and then postures are recognized by using the well known K-nearest neighbor and support vector machine classifiers. It has been shown that for posture recognition, the features obtained from FD perform better than that obtained from the Hu moments [10], [8]. In [6], it is reported that the orthogonal 2D Zernike moments (ZMs) as features of postures provide significantly better recognition accuracy than the Hu moments do. It may be found that to construct the features of the hand postures, the ZM-based recognition method as well as the methods based on other orthogonal polynomials (e.g., Tchebichef and Krawtchouk), chooses the order of moments heuristically without any justification [6]–[8].

We argue that only the orthogonal moments those have significant discrimination power may be used to construct the feature vector instead of choosing moments of an arbitrary order. In this paper, the complex orthogonal ZMs are chosen for features of posture due to the fact that they show high noise resilience, significantly low information redundancy, and efficient reconstruction capability in image analysis [11], when comparing with other orthogonal moments. In this context, the orthogonal 2D ZMs only those have high value of interclass to intraclass scatter ratio are used to construct the features of hand postures. Experiments are conducted to show the significance of such approach of constructing the features on the posture recognition accuracy.

The paper is organized as follows. In Section II, a brief review of the orthogonal 2D moments based on the Zernike polynomials is given. The proposed posture recognition method using the discriminative ZMs (DZMs) is presented in Section III. In Section IV, experimental results of recognition performance on commonly-referred hand posture database are given. Finally, concluding remarks are provided in Section V.

II. 2D Zernike Moments: A Brief Review

In 1934, Zernike proposed a set of complex orthogonal polynomials defined on the unit disc $(x^2 + y^2) \leq 1$, where (x, y) is a spatial location on the 2D Cartesian coordinates [12]. In the literature, these polynomials are known as the Zernike polynomials. The complex orthogonal moments obtained from the projections of the basis functions those are constructed from the Zernike polynomials are known as 2D Zernike moments [13]. Let $f(r, \theta)$ be the integrable image intensity function representing the still hand posture in polar coordinates considering the center of the image as the origin. The complex 2D Zernike moments of order n $(n \in 0, 1, 2, \cdots, \infty)$ and repetition ℓ $(\ell \in -\infty, \cdots, -2, -1, 0, 1, 2, \cdots, \infty)$ of the postures are defined as [13]

$$M_{n\ell} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f(r,\theta) Z_{n\ell}(r,\theta)^* r \mathrm{d}r \mathrm{d}\theta \tag{1}$$

where $*$ denotes the complex conjugate, $|\ell| \leq n$, and $n - |\ell|$ is even. The complex Zernike polynomials in a unit disk of polar coordinates are defined as [13]

$$Z_{n\ell}(r,\theta) = R_{n\ell}(r) e^{in\ell\theta} \tag{2}$$

where $i = \sqrt{-1}$ and $R_{n\ell}(\cdot)$ is real-valued orthogonal radial polynomial defined as [13]

$$R_{n\ell}(r) = \sum_{m=0}^{\frac{n-|\ell|}{2}} (-1)^m \frac{(n-m)!}{m! \left(\frac{n+|\ell|}{2} - m\right)! \left(\frac{n-|\ell|}{2} - m\right)!} r^{n-2m} \tag{3}$$

The posture image may be reconstructed from the estimated 2D complex ZMs as

$$\hat{f}(r,\theta) = \sum_{n=0}^{\eta} \sum_{\ell} M_{n\ell} Z_{n\ell}(r,\theta) \tag{4}$$

where η is the maximum order of moments. In order to improve the readability, the subscripts $n\ell$ will be replaced by ν in the remainder of the paper to denote each of Γ number of ZMs given by

$$\Gamma = \begin{cases} \frac{1}{4}\left[(n+2)^2 - 1\right] & \text{if } n \text{ is odd and } \ell \geq 0 \\ \frac{1}{4}\left[n+2\right]^2 & \text{if } n \text{ is even and } \ell \geq 0 \end{cases} \tag{5}$$

III. Discriminative Zernike Moments for Hand Poster Recognition

In this section, we consider a hand posture recognition algorithm such that postures of class label k are stored in terms of the ZMs denoted as M_ν^k $(\nu \in 0, 1, 2, \cdots, \Gamma)$ $(k \in 1, 2, \cdots, N)$ in a database, where total number of posture class is N. It is noted that predefined linear combinations of these moments may form geometric moments those are invariant to certain distortions. In the proposed method, orthogonal ZMs are treated as independent features due to the fact that no specific geometric distortion is ensured to exist in the hand posture database. In order to construct the ZM-based feature sets for the purpose of posture recognition, only those moments that have high discrimination capability are selected. In this regard, we compute the scatter ratio of interclass to intraclass variances for each of the moments individually to estimate their independent discrimination capabilities, and latter construct the moment-based features for hand posture recognition using the discrimination powers.

Let λ_{tr}^k be the number of training sets of complex ZMs for certain posture class k $(k \in 1, 2, \cdots, N)$. In order to calculate the discrimination capability of each of the moments, we have calculated the contribution of the variance of the moment within a given class with respect to the total variance of the database. The average of within class variance V_ν^C of a complex moment is given by

$$V_\nu^C = \frac{1}{N} \sum_{k=1}^{N} \sum_{s=1}^{\lambda_{tr}^k} \left| M_\nu^k(s) - \mu_\nu^k \right|^2 \quad \nu = 0, 1, 2, \cdots, \Gamma \tag{6}$$

where $|\cdot|$ represents the absolute value of a complex moment and μ_ν^k is the mean of the moment within the class k $(k \in 1, 2, \cdots, N)$ given by

$$\mu_\nu^k = \frac{1}{\lambda_{tr}^k} \sum_{s=1}^{\lambda_{tr}^k} M_\nu^k(s) \quad \nu = 0, 1, 2, \cdots, \Gamma \tag{7}$$

In a similar fashion, the total variance V_ν^T of the moment may be obtained as

$$V_\nu^T = \sum_{k=1}^{N} \sum_{s=1}^{\lambda_{tr}^k} \left| M_\nu^k(s) - \mu_\nu \right|^2 \quad \nu = 0, 1, 2, \cdots, \Gamma \tag{8}$$

where μ_ν is the mean of the moment in the entire training database given by

$$\mu_\nu = \frac{1}{\lambda_{tr}^k N} \sum_{k=1}^{N} \sum_{s=1}^{\lambda_{tr}^k} M_\nu^k(s) \quad \nu = 0, 1, 2, \cdots, \Gamma \tag{9}$$

In the proposed recognition method, the discrimination power D_ν of ZM is defined as

$$D_\nu = \frac{V_\nu^T}{V_\nu^C} \quad \nu = 0, 1, 2, \cdots, \Gamma \tag{10}$$

A high value of D_ν implies that the complex moment M_ν possesses a low within class variability with respect to the total variability. In other words, a high value of D_ν indicates a high discrimination capability of the corresponding moment. Thus, it is expected to attain the maximum recognition rate by constructing the feature vector of posture images such that the feature vectors comprise only those ZMs that have significantly high discrimination power.

Hence, we select as features only those ZMs that correspond to the α $(\alpha \ll \Gamma)$ largest values of D_ν, α being the number of moments selected for classification. In such a

TABLE I. RESULTS CONCERNING THE RECOGNITION ACCURACY OF HAND POSTURES CONSIDERED IN THE EXPERIMENTS

Experimental Methods	Accuracy in Percentage				Number of Features
	$\lambda_{tr}=4$	$\lambda_{tr}=6$	$\lambda_{tr}=8$	$\lambda_{tr}=10$	
PCA [1]	76.61	89.26	93.96	96.65	λ_{tr}
FD [5]	90.11	93.33	95.03	95.68	20
ZM [6]	90.95	93.64	95.33	96.37	34
Proposed DZM	**96.16**	**97.30**	**97.69**	**98.51**	34

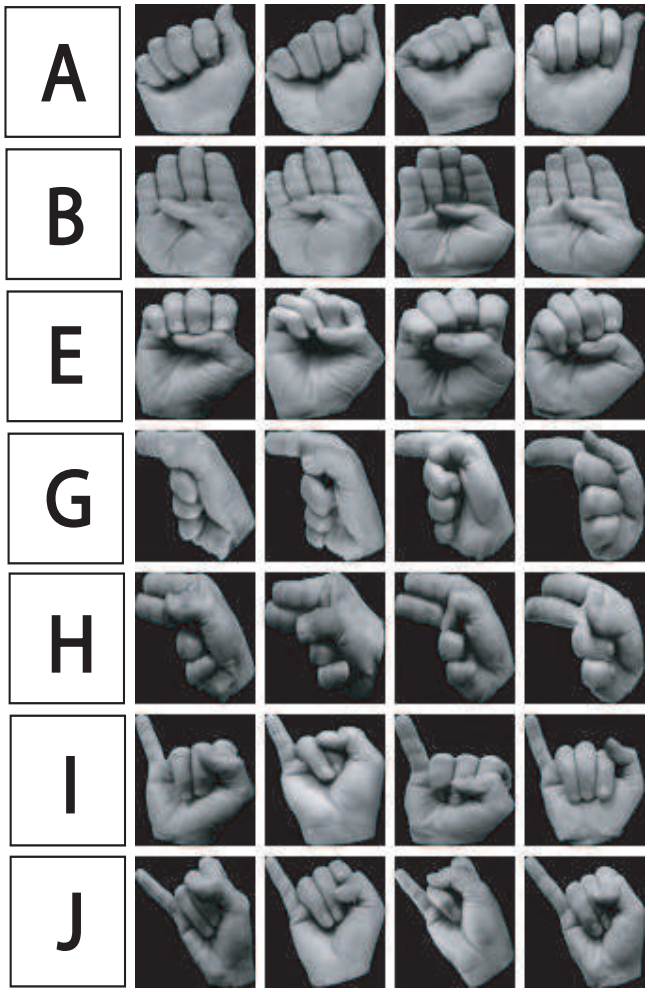

Fig. 1. Four samples for each letter of English alphabets, viz., 'A', 'B', 'E', 'G', 'H', 'I' and 'J' in the MUGD database as per the guidelines of the ASL.

case, the feature vector for the s-th image of class label k $(k \in 1, 2, \cdots, N)$ is given by

$$\mathbf{\Phi}_s^k = \left[\phi_{1s}^k, \phi_{2s}^k, \phi_{3s}^k, \cdots, \phi_{\alpha s}^k\right]^T \tag{11}$$

where T denotes the transpose of the vector and $\phi_{rs}^k \in \{M_\nu^k(s) : \nu = 0, 1, 2, \cdots, \Gamma\}$ is the complex moment corresponding to the r-th element of the vector

$$\mathbf{D}_{\text{sort}} = \left[D_1, D_2, \cdots, D_\alpha, D_{\alpha+1}, \cdots, D_\Gamma\right]^T \tag{12}$$

which comprises the discriminative powers arranged in descending order. Note that to construct the feature vector $\mathbf{\Phi}$ for a posture image an important parameter is the α that determines the length of the feature set. In the proposed method, we select only those moments that have within class variance at least τ $(\tau > 1)$ times lower than the total variance of the training database. In other words, the parameter α is chosen in such a way that

$$D_\alpha \geq \tau \tag{13}$$

In practice, the number of image samples available for each class are comparable with the number of total classes of hand postures, and hence, D_α provides a good clustering performance. In the proposed method, with a view to recognize the posture image in test in a computationally efficient way, the complex feature $\mathbf{\Phi}$ is used in the well-known non-parametric classification algorithm, viz., the nearest neighbor.

IV. EXPERIMENTAL RESULTS

Extensive experimentations are carried out to investigate the performance on recognition accuracy of the hand posture images using the proposed DZM method as compared to the existing methods. In this paper, we present the results that are obtained using one of the most exhaustive hand gesture databases, viz., the Massey University Gesture Dataset (MUGD) 2012 [14]. The database consists of 26 static gesture of English alphabet and 10 numeric gestures as per the guidelines of the American Sign Language (ASL). There are 2425 color images of hand postures obtained from five individuals. The images have notable variations in terms of illumination, scaling, as well as rotation both in-plane and out-of-plane. Since the posture images in the database are of different sizes, all the images are normalized to pixel size of 64×64. Although experiments are carried out for all the 36 classes of hand postures, as representative results, only the recognition performance of 10 alphabetical gestures, namely, 'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I' and 'J', are presented in this paper. Fig. 1 shows four samples of grayscale hand postures of representative seven letters in the database. It can be seen from this figure that the samples of gestures of 'A' have both the scaling and in-plane rotations and that of 'B' have both the out-of-plane rotations and illumination variations. It is noted that the samples of gestures of 'I' and 'J' are very challenging to recognize, especially since one is approximately the rotated version of the other.

Four hand posture recognition methods, viz., the PCA [1], FD [5], existing ZM [6] and the proposed DZM-based methods are considered in the experiments. In the case of PCA-based method, the number of eigenvectors for generating feature set is chosen as 10% of the total number of training images, since such a choice provides an optimum recognition

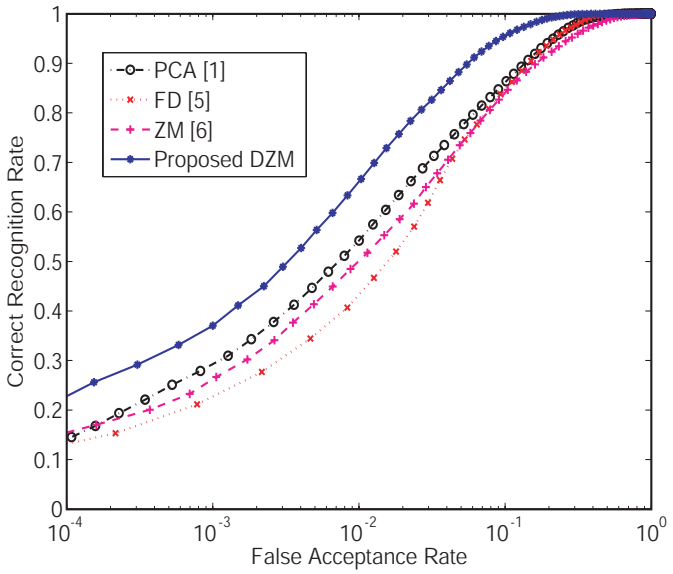

Fig. 2. Receiver operating characteristics curves showing the comparison of correct recognition rates as a function of false alarm rates obtained from the methods considered in the experiments when $\lambda_{tr} = 10$.

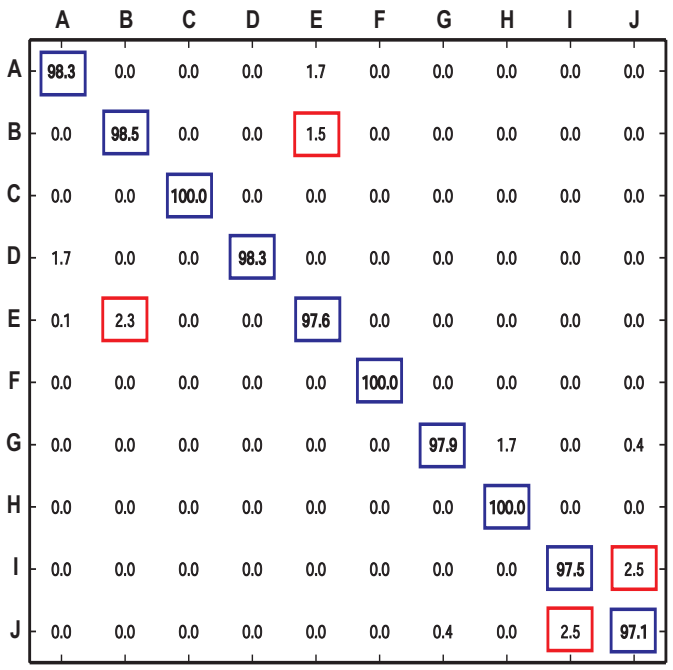

Fig. 3. Confusion table showing the classification results provided by the proposed DZM method considering $\lambda_{tr} = 10$.

performance [15]. The number of points in the boundary of the posture shapes are taken as 20 for the FD-based method as recommended in [5]. The existing ZM-based method employed the nearest neighbor distance of the feature vector considering the magnitudes of the entire set of moments up to order 10 except the first two [6]. In the case of the proposed method, the discriminative complex moment-based features are constructed by choosing a value of τ in such a way that the number of moments equal to that of [6]. These moments are selected from an entire set of moments those are obtained from choosing the highest order ZM to be 15. In order for the results to be statistically robust in terms of accuracy, λ_{tr} number of training postures are chosen randomly from the database and the rest as the test postures. The recognition accuracy is measured as the percentage of test postures those classified accurately. The results presented in the paper consider that such recognition accuracies are averaged over 20 random sets of training images.

Table I shows the recognition accuracies in percentage obtained from the methods considered in the experiments for varying number of training posture images. This table also shows the number of features used for each of the methods. It can be seen from this table that the recognition accuracies increase with the number of training images for all the methods. The results shown in the table reveal that in general ZMs are effective for recognizing the hand postures as compared to the PCA [1] or FD-based [5] methods. Further, the proposed DZM method that selects only the discriminant moments for construction of the features provides more than 3% recognition accuracy on average as compared to the existing ZM-based method [6]. Since the DZM and ZM-based methods use same number of features, the execution period for obtaining the decision on the hand postures is the same for both these methods. The proposed method requires additional computational load only in the enrollment process to find the discriminant moments.

Fig. 2 shows the receiver operating characteristics (ROC) curves showing the comparisons of correct recognition rate of the class of hand postures as a function of false alarm rate obtained from the methods considered in the experiments when $\lambda_{tr} = 10$. The results in terms of ROC curves reveal that the probability of correct recognition provided by the proposed DZM method is always the highest for any false alarm rate as compared to that provided by the other methods. Fig. 3 shows the confusion table depicting the classification results obtained by the proposed DZM method using $\lambda_{tr} = 10$. From this figure, it can be seen that even for this small number of training posture images, the proposed method is capable of providing a recognition accuracy of more than 97% for all cases. The challenges arise in the recognition of two pairs, viz., 'B and E' and 'I and J' due to the fact that the postures for these pairs appear to be very close to each other [14]. In the experiments, it is found that such an error does not exceed more than 2.5% in the case of the proposed DZM method.

V. Conclusion

This paper presents a novel static hand gesture recognition algorithm, which has many potential applications related to vision-based communication in the area of HCI. Orthogonal 2D ZMs have been chosen to construct the features of posture images due to their well known properties of invariance towards the rotation, scaling and shifting of the patterns. The key contribution of the paper lies in the fact that instead of selecting a heuristic set of higher-order ZMs, the posture features have been obtained using only those moments possessing significant amount of discrimination capabilities. Due to the orthogonal nature of the moments, the discrimination capabilities of the moments have been evaluated individually using the ratio of inter- and intraclass variances. Experiments carried on commonly-referred hand posture database have shown that

the proposed DZM method, which selects the discriminative moments in constructing the features results in significant improvement in the recognition accuracy as compared to the existing methods. As a further work, the investigations can be conducted to improve the recognition accuracies of the hand postures those have very similar patterns.

REFERENCES

[1] H. Birk, T. B. Moeslund, and C. B. Madsen, "Real-time recognition of hand alphabet gestures using principal component analysis," in *Proc. 10th Scandinavian Conference on Image Analysis*, Lappeenranta, Finland, 1997, pp. 261–268.

[2] S. C. W. Ong and S. Ranganath, "Automatic sign language analysis: A survey and the future beyond lexical meaning," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 873–891, 2005.

[3] N. Boughnim, J. Marot, C. Fossati, and S. Bourennane, "Hand posture classification by means of a new contour signature," in *Lecture Notes in Computer Science : Advanced Concepts for Intelligent Vision Systems*, Brno, Czech Republic, 2012, pp. 384–394.

[4] P. Trigueiros, F. Ribeiro, and L. P. Reis, "A comparative study of different image features for hand gesture machine learning," in *Proc. 5th Int. Conf. Agents and Artificial Intelligence*, Barcelona, Spain, 2013, pp. 51–61.

[5] C. W. Ng and S. Ranganath, "Real-time gesture recognition system and application," *Image and Vision Computing*, vol. 20, no. 13-14, pp. 993–1007, 2002.

[6] K. C. O.-Rodriguez, G. C.-Chavez, and D. Menotti, "Hu and Zernike moments for sign language recognition," in *Proc. Int. Conf. Image Processing, Computer Vision, and Pattern Recognition*, Las Vegas, NV, 2012, pp. 1–5.

[7] S. P. Priyal and P. K. Bora, "A study on static hand gesture recognition using moments," in *Proc. Int. Conf. Signal Processing and Communications*, Bangalore, India, 2010, pp. 1–5.

[8] S. Bourennane and C. Fossati, "Comparison of shape descriptors for hand posture recognition in video," *Signal, Image and Video Processing*, vol. 6, no. 1, pp. 147–157, 2012.

[9] J. Rekha, J. Bhattacharya, and S. Majumder, "Hand gesture recognition for sign language: A new hybrid approach," in *Proc. Int. Conf. Image Processing, Computer Vision, and Pattern Recognition*, Las Vegas, NV, 2011, pp. 80–86.

[10] S. Conseil, S. Bourennane, and L. Martin, "Comparison of Fourier descriptors and Hu moments for hand posture recognition," in *Proc. European Signal Processing Conference*, Poznan, Poland, 2007, pp. 1960–1964.

[11] C. Teh and R. T. Chin, "On image analysis by the method of moments," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 496–513, 1988.

[12] F. Zernike, "Diffraction theory of the cut procedure and its improved form, the phase contrast method," *Physica*, vol. 1, pp. 689–704, 1934.

[13] M. R. Teague, "Image analysis via the general theory of moments," *J. Optical Society of America*, vol. 70, no. 8, pp. 920–930, 1979.

[14] A. L. C. Barczak, N. H. Reyes, M. Abastillas, A. Piccio, and T. Susnjak, "A new 2D static hand gesture colour image dataset for ASL gestures," *Research Lett. Information and Mathematical Sciences*, vol. 15, pp. 12–20, 2011.

[15] P. K. Pandey, Y. Singh, and S. Tripathi, "Image processing using principle component analysis," *Int. J. Computer Applications*, vol. 15, no. 4, pp. 37–40, 2011.