

Recognition of static gestures using correlation and cross-correlation



Shazia Saqib *, Syed Asad Raza Kazmi

Department of Computer Science, GC University, Lahore, Pakistan

ARTICLE INFO

Article history:

Received 22 December 2017

Received in revised form

16 March 2018

Accepted 24 March 2018

Keywords:

Sign language

Correlation

Cross correlation

Invariance

Norm correlation

Normalized cross correlation

RGB

ABSTRACT

Sign language recognition has been an active area of research for around two decades and numerous sign languages have been extensively studied in order to design reliable sign language recognition systems. Pakistan sign language (PSL) has been used as a case study here. A comprehensive database of static images depicting the signs for different Urdu alphabets is being used as a reference and input images are being compared to perform PSL alphabet recognition. The normalized Correlation technique is being used for image registration between input image and images from the database to find the closest match. The purpose of research is to identify static gestures of any Sign Language which can ultimately lead to an identification of words and sentences. The research starts with image acquisition, image preprocessing, and use of correlation and labeling of an identified symbol. Normalized correlation is used to find the nearest match. This paper includes experiments for 37 static hand gestures related to PSL alphabets. Training dataset consists of 10 samples of each PSL symbol in different lighting conditions, different sizes and shapes of hand by 5 different signers. This gesture recognition system can identify one hand static gestures in any complex background with a “minimum-possible constraints” approach. A comparison is also drawn between normalized correlations and normalized cross-correlation. As compared to other technique, this technique can work with a small dataset size. The technique is based on unsupervised learning.

© 2018 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

In our day to day life we deal with huge masses of digital visual information. Any image processing system uses image analysis techniques as major requisite. Need is arising to analyze the embedded visual information. Gestures provide an attractive, user-friendly alternative to using an interface device like a keyboard, mouse, and joystick in human-computer interaction (HCI). Every gesture recognition research targets to build a system that can identify the information passed through human gestures automatically and to make use of them to convey information (i.e., for communicative use as in sign-language) or to control devices (i.e., manipulative use as in controlling robots without any physical contact). Unfortunately, in this country people who have an in-depth knowledge of sign language are very rare, which has led to the social seclusion of the deaf community. There are two

kinds of gestures; static gestures and those which include dynamic hand, body and face movements (Saqib and Kazmi, 2017). For static gestures, within a particular time-frame, the noticeable gesture is physically coordinated. For dynamic gestures, series of finger and hand poses are recognized and analyzed

This paper can be divided into six sections. Introduction is followed by Section II, which provides a detailed overview of related work in different systems dealing with sign language recognition. Section III focuses on proposed approach and it briefs on the salient features of the algorithm used to distinguish between different gestures and provides a step-by-step procedure within the proposed algorithm. Section IV elaborates experimental setup, Section V discusses the results and Section VI concludes the major findings of this paper and provides guideline for future work.

2. Literature review

Within the past two decades, comprehensive literature has been accumulated related to the numerous geographically global and local sign languages as well as the different methods employed

* Corresponding Author.

Email Address: shaziasaqib@lgu.edu.pk (S. Saqib)

<https://doi.org/10.21833/ijaas.2018.06.002>

2313-626X/© 2018 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

to resolve the issue of sign language detection through image-based models.

Jie et al. (2016) recognized real-time continuous gesture recognition of sign language using a DataGloveTM. Hidden Markov models (HMMs) is used for identification of 51 fundamental postures, 6 orientations, and 8 motion primitives with the average recognition rate at 80.4%. Alvi et al. (2007) has presented recognition of PSL gestures using statistical template matching using Data Glove developed by 5DT in his project (i.e., Boltay Haath). Use of camera for recognition, which being affected by environmental changes, does not give good results. The system recognizes one handed alphabet signs from PSL.

Nagi et al. (2008) have presented a new technique for human face recognition. The technique helps in recognition of faces; it removes redundant details and reduces them in size by using 2D cosine transform (2D-DCT). The technique uses skin color, and DCT coefficients. It uses an unsupervised learning technique to classify DCT-based feature vectors into groups to identify if the input image matches some image in dataset or not.

In 2008, Kausar et al. (2008) have used, a fuzzy classifier to recognize alphabets of Pakistani sign language however her technique is dependent on the use of color gloves. The color gloves help to identify each finger-tip and joint using the angles between fingers help in identification of the alphabet of sign language. She got highly accurate results except 2, all other 35 signs have been recognized accurately. Sarkalehl et al. (2009) suggested the use of discrete wavelet transform (DWT), and then application of a multi layered Perceptron (MLP) Neural Network (NN) to classify the selected images. This technique does not use any gloves or visual marking systems. It gives an accuracy of 98.75% when the network is trained using MATLAB NN Toolbox. Tauseef et al. (2009) developed a scheme for static gestures. The success of the approach is evident from the results and we achieved an accuracy rate of 97.4%. The approach is affected by the number of regions produced during color segmentation of the image which may cause a decrease in accuracy rate in some situation.

Shahzad et al. (2009) presented an online system for recognizing isolated, hand-sketched Urdu characters drawn on a Tablet PC. Attributes of Urdu characters are analyzed to define a set of features which are then trained and classified using a weighted, linear classifier. Moghaddam et al. (2011) used kernel based feature extraction methods kernel principle component analysis (KPCA) and kernel discriminant analysis (KDA) on Persian sign language (PSL) postures. To compare the impact of features on signs' recognition rate, classifiers such as minimum distance, support vector machine (SVM) and Neural network (NN) is used. Jalilian and Chalechale (2014) made use of feed forward neural networks and recurrent neural networks along with its different architectures; partially and fully recurrent networks. The system works well for static

gesture recognition at an accuracy of 95%. Parveen et al. (2011) used Conditional Random Fields for classification of clause boundary beginning and ending and also detecting the type of subordinate clause. Limitation with CRFs is that it is highly dependent on linguistic rules. Missing of these rules may lead to wrongly classified data. Bansal et al. (2011) have used Hidden Markov Model as an indispensable tool for the recognition of dynamic gestures in real time. They have suggested standardizing the axis through the centroid thus greatly reducing the database size.

Albelwi and Alginahi (2012) used a vision-based automatic sign language recognition system for Arabic letter using Haar classifiers. The skin detection is done in HSV color space. Transforming images into frequency domain using Fourier Transform gives entire signs database into a single vector based on FDs to ease the matching process thus giving upto 90.55 % recognition accuracy in real time. Naoum et al. (2012) presented a new Arabic sign language recognition using K-nearest Neighbor algorithm. K-Nearest Neighbor Algorithm and feature extraction are the guidelines of the recognition system, because hand gestures is treated as a block of curves needed to be extracted in the best fit with a predefined character set in the knowledge base. Sasirekha and Chandra (2012) focused on the PDF images, but variations in text style, font, size, orientation, alignment make the problem an uphill task. This paper suggests two techniques under block-based classification.

Pansare et al. (2012) presented use of filters, cropping using a feature vector as centroid and area of edge, which was compared with feature vectors of a training dataset of gestures using Euclidian distance in the fourth stage. Least Euclidian distance gives recognition of perfect matching gesture for display of ASL alphabet, meaningful words using file handling. Abdalla and Hemayed (2013) suggested the use of hand blob using YCbCr color space to detect skin color of hand. The system classifies the input pattern based on correlation coefficients matching technique. The experiment results show that the gesture recognition rate of 20 different signs, performed by 8 different signers, is 85.67%. M. Nachamai (2013) has tried to approach this in a simple but efficient manner using the basic SIFT algorithm for recognition. The efficacy of the approach is proved well through the results obtained, invariably on both the datasets.

Ali (2013) has worked on Urdu sign language. He uses text to sign and sign to text conversion. First takes an input text from a text box and shows image on output display. Secondly, it takes a sign through webcam and outputs the equivalent text. Singha and Das (2013) made use of Eigen value weighted Euclidean distance for classification of various Sign Languages of India. Skin Filtering, Hand Cropping, Feature Extraction and Classification are four stages of the process. 24 signs with 10 samples each were considered for which recognition rate obtained was 97%.

Szegedy et al. (2013) have done object detection using DNNs. The approach presented use of regression problem to object bounding box masks. A multi-scale inference procedure is used to produce high-resolution object detections at a low cost by a few network applications. Sykora et al. (2014) suggested two feature extraction methods: SIFT as the first method and SURF method as second. They were applied on set of depth map images of left hand gestures. There were 10 gestures. For capturing these images the Microsoft Kinect camera was used. For image classification the Support vector machine was used. The experimental results are prediction accuracies of SVM method on test set images for each descriptor. Sahoo et al. (2014) used two data sets which contains 2600 images for single handed characters and 2340 gestures of double handed characters (A-Z). The structural features, local histogram features and direct pixel values of gray scale images extracted from these gestures are used as input to the recognition system. After extracting features from images, kNN classifier and neural network classifier are used to classify the gestures. In single handed data set 95.30% recognition rates are achieved, and in double handed data set 96.37% accuracy rates are achieved. Mohandes et al. (2014) have combined the enhanced skin, motion, and depth feature in particle filter model, the performing hand can be well localized and tracked in every frame. A shape-order context descriptor is then proposed for gesture sequence matching in temporal spatial domain. Such a rich descriptor can greatly improve the gesture recognition rate and be invariant to gesture to translate and scale. Pandey and Jain (2015) have used Cross-correlation technique for image registration between input image and images from the database to find the closest match. The tolerance level ensures a trade -off between computational complexity and accuracy of the match between set of images. The algorithm tests on the images have been around 75% successful and attempts are being made for more efficient and robust performance.

Bhuyan et al. (2014) entailed using gesture spotting to distinguish meaningful gestures from unintentional movements. To avoid the effects of variations in a gesture's motion chain code (MCC), the orientation and length of an ellipse least-squares fitted to motion-trajectory points and the position of the hand are used. The recognition rate for static gestures is almost 96.0 % using the proposed features and for continuous gestures, the recognition rate for the proposed features was 88.9 %. Ohn-Bar and Trivedi (2014) have developed a vision-based system that employs a combined RGB and depth descriptor to classify hand gestures. The feasibility of the system is demonstrated using a challenging RGB hand gesture data set collected under settings of common illumination variation and occlusion. Fernandez et al. (2015) proposed and investigated presents a solution in the form of ZACFs and RACF that completely remove circular correlation effects from CF designs. To address the computational

challenges caused by the ZA constraints we introduced the RACF designs as an approximate solution as well as proximal gradient descent based algorithms for exactly solving for the various ZACFs.

Zhou et al. (2014) have used large datasets like ImageNet and the rise of Convolutional Neural Networks (CNNs) for learning high-level features. Performance at scene recognition has not attained the same level of success. This may be because current deep features trained from ImageNet are not competitive enough for such tasks. New methods to compare the density and diversity of image datasets and show that places are as dense as other scene datasets and has more diversity.

Shurong et al. (2015) have proposed that the dynamic sign language can be described by a sequence of key frames and then recognized by these key frames. This method focuses on the key frame extraction and can minimize the limitation to the users and the requests of equipment, which makes the interaction between a human and computer more natural and realizes the comprehensive application of sign language recognition. Abdo et al. (2015) have introduced an Arabic Alphabet and Numbers Sign Language Recognition (ArANSLR). The technique recognizes the alphabet and numbers signs of Arabic sign language to text or speech. Experiments on real-world datasets showed that the proposed algorithm for Arabic alphabet and numbers sign language recognition is suitability and reliability compared with other competitive algorithms.

Tehsin et al. (2013) have compared different shape descriptors. She has used two groups of signs to analyze, for central moments and Hu Moments, whereas Fourier descriptor have given have given very encouraging results for both the groups. Raees et al. (2016) suggested a deep pixels-based analysis for the recognition of fingers (from index to small finger) while thumb position is determined through Template Matching. The system's accuracy achieved a satisfactory level of 84.2% when evaluated with signs comprising 180 digits and 240 alphabets.

Darwish et al. (2016) have presented several results concerning static hand gesture recognition using an algorithm based on Type-2 Fuzzy HMM (T2FHMM). The features used as observables in the training as well as in the recognition phases are based on Singular Value Decomposition (SVD) that optimally exposes the geometric structure of a matrix. SVD is an extension of Eigen decomposition to suit non-square matrices to reduce multi-attribute hand gesture data to feature vectors. The recognition rate of the proposed system is 100% for uniform hand images and 95.5% for cluttered hand images. Jain (2016) has designed and developed an application Zauber. The windows based application is able to recognize gestures and act like a remote to windows media player at a high rate using Haar-like feature for gesture detection. Singha and Laskar (2017) have introduced a two-level speed normalization procedure using DTW and Euclidean distance-based techniques. Three features such as

'orientation between consecutive points', 'speed' and 'orientation between first and every trajectory points' were used for the speed normalization. An accuracy of 94.78 % was achieved using the classifier fusion technique as compared to baseline CRF (85.07 %) and HCRF (89.91 %) models. Zhang et al. (2017) used fully end-to-end approach for visual tracking in videos that learns to predict the bounding box locations of a target object at every frame. Based on this intuition, he has formulated his model as a recurrent convolutional neural network agent that interacts with a video overtime, and the model can be trained with reinforcement learning (RL) algorithms to learn good tracking policies that pay attention to continuous, inter-frame correlation and maximize tracking performance in the long run.

3. Proposed approach

In an image recognition system, the process starts with image acquisition and ends at recognition of the image. A large number of techniques exist that are grouped into supervised learning, unsupervised learning, reinforcement learning and deep learning. The recognition techniques usually use algorithms in the domains of Logistic Regression, Decision Tree, SVM, Naive Bayes, KNN, K-Means, Random Forest, Dimensionality Reduction Algorithms, and Gradient Boosting Algorithms. This paper suggests use of correlation in combination with filters and morphological operations.

3.1. Difficulties in object recognition under varied circumstances

Any sign recognition system faces following difficulties:

3.1.1. Lightning

The lightning conditions may differ during the course of the day and also from region to region. Change in weather also changes the lighting in an image. In-door and outdoor images for same object can have varying lightning condition. Shadows in the image can affect the image light. Whatever the lightning may be the system must be able to recognize the object in any of the image. Fig. 1 shows different problems encountered in gesture recognition system.

3.1.2. Positioning

Position in the image of the object can be changed. If template matching is used, the system must handle such images uniformly.

3.1.3. Rotation

The image can be in rotated form. The system must be capable to handle such difficulty. As shown in Fig. 2, the character "ālf" can appear in any of the

form. But the orientation of the letter must not affect the recognition of the character or any image.

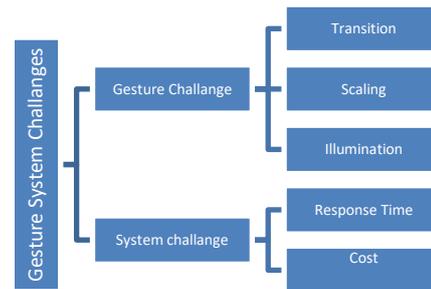


Fig. 1: Challenges in gesture recognition system



Fig. 2: Different orientations of alphabet "ālf"

3.1.4. Mirroring

Any good recognition system must take into care the mirror images as well.

3.1.5. Obstacle

This is the situation when object is not visible rather it is obscured by some other object. The system of object recognition must handle such type of condition and in the output the correct result.

3.1.6. Scale

The system should be able to handle any variance in the size of the object.

Any object recognition scheme must handle all above problems. An efficient and robust object detection system can be developed by conquering the above stated difficulties.

3.2. Approaches for object recognition

3.2.1. Template matching

Template matching is a technique for finding small parts of an image which match a template image. We match the input image with the images in data set. Templates help identifying characters, numbers, objects, etc. It can be performed on either color or gray level images. Template matching can either be pixel to pixel matching or feature based.

3.2.2. Color based

Another approach to compare images is colored based using histograms. These techniques use the incorporation of color for object detection based on the above mentioned criteria and demonstrate the

advantages of combining color with shape. We get encouraging results on detection of gestures using even challenging datasets. The color models give good response to object occlusion and cluttering; robustness to noise in the images.

3.2.3. Shape based

New techniques in object recognition are using shape based features to identify objects. Shape features are often used as a replacement or complement to local features.

3.3. Suggested solution

This paper presents a very simple technique for object recognition based on correlation. The approach suggests use of cross-correlation to detect and recognize images. Fig. 3 shows steps involved in the image recognition system.

Correlation is used to test relationship between two things. Usually in place of simple correlation, normalized correlation is used. The value of cross-correlation value varies between 1 to -1. A 0 value of normalized correlation means no relation at all. But a positive or negative value of correlation gives measure of the way two things are related.

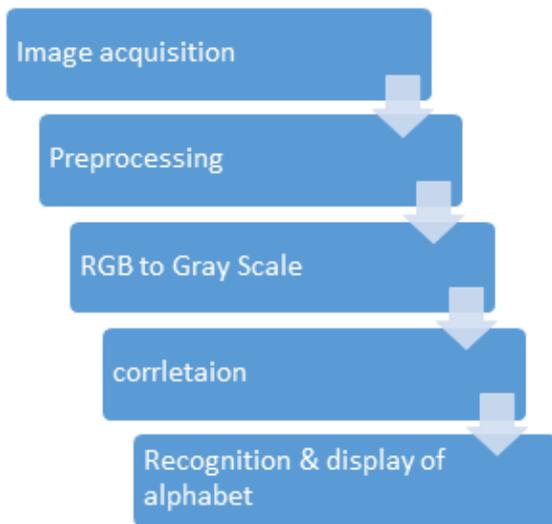


Fig. 3: Process of gesture recognition

Correlation when applied to images gives the similarity between the two. It is basically sum of products of the two vectors whose correlation we are determining.

$$\text{corr}_{x,y} = \sum_{n=0}^{n-1} x[n] \cdot y[n] \tag{1}$$

However normalized correlation gives a value between -1 and 1.

$$\text{corr - norm}_{x,y} = \frac{\sum_{n=0}^{N-1} x[n] \cdot y[n]}{\sqrt{\sum_{n=0}^{n-1} x[n]^2 \cdot \sum_{n=0}^{n-1} y[n]^2}} \tag{2}$$

The normalized correlation value r can help to compare similarity of two different vectors.

$$r = \pm 0.8 \text{ or higher} \Rightarrow \text{strong correlation} \tag{3}$$

$$r = \pm 0.5 \text{ .. } \pm 0.8 \Rightarrow \text{medium correlation} \tag{4}$$

$$r = \pm 0.4 \text{ or lower} \Rightarrow \text{weak correlation} \tag{5}$$

4. Experimental setup

The experimental setup starts with capturing image with the help of laptop camera or mobile camera. The input images are then compared with the images in the dataset after preprocessing phase. The closest match is the image input.

4.1. Preprocessing

Image preprocessing in all image recognition systems includes the set of operations on images whose goal is the improvement of the image data that suppresses undesired distortions or enhances some image features important for further processing. Hand region is extracted from the image using various image processing techniques like image cropping, background subtraction, noise removal using any median or Gaussian filter etc. The image is structured as a flat rectangular shape. The algorithm for the entire process is as under:

1. Input the image A.
2. Reduce the resolution of A by a factor of 10 to increases speed of future calculations. Remove noise using any box filter or Gaussian Filter.
3. Convert to Gray scale from RGB A_{grey}
4. Compare input image A_{grey} with the images in the dataset one by one using normalized correlation.
5. Find the highest or lowest value of normalized correlation r. If $|r| > 0.75$, then we declare the image as recognized.

Table 1 shows that for 37 alphabets, 370 images were tested and we got incorrect result for 40 images. This gives an accuracy of 89.2%.

Table 1: detail of images tested

Total alphabets	37	370
Correct detection	34	330
False detection	3	40

Fig 4, 5, and 6 give correlation of “ālf”, “ṭ”, “ḍ” with other Urdu language alphabets. “ālf” and “ṭ” are recognized but for “ḍ” are not recognized correctly.

However system works good when there is only one signer, for more than one signer correlation fails to give high accuracy in recognition. There are many environments where device needs only one signer. In those scenarios we can use correlation for image recognition.

4.2. Comparison with cross-correlation

Normalized Cross Correlation (NCC) is used to compare an input image with a bigger image. It is usually used to find one image in another image. It is also affected by nonuniform illumination and shadows. It calculates the degree to which two

signals resemble. It has been commonly used as a metric to evaluate the degree of similarity (or dissimilarity) between two compared images. The main advantage of the normalized cross correlation over the ordinary cross correlation is that it is less sensitive to linear changes in the amplitude of illumination in the two compared images. Like simple normalized correlation, it also returns value in the range between -1 and 1.

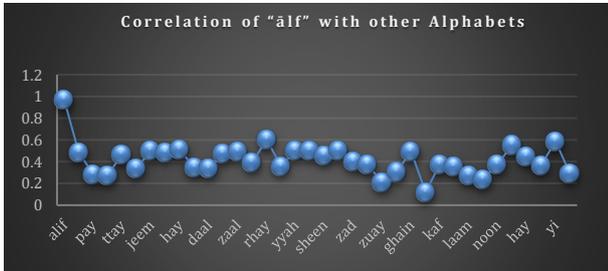


Fig. 4: Normalized correlation of “ālf” with other alphabets

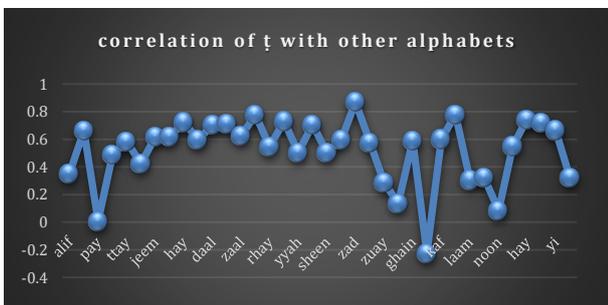


Fig. 5: Normalized correlation of “ṭ” with other alphabets

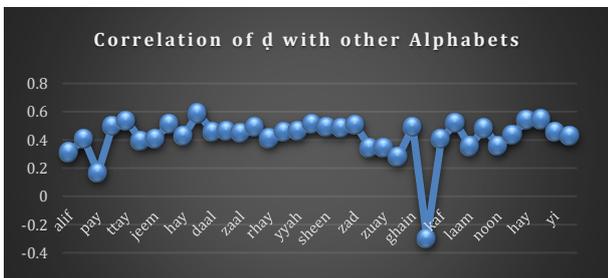


Fig. 6: Normalized correlation of “ḍ” with other alphabets

Calculations in normalized cross correlation are more complex as compared to simple normalized correlation. Fig. 7 shows comparison of input image with images in data set. The green rectangle shows the starting point of similarity of both the images.



Fig. 7: Use of cross correlation in image matching

The following Eq. 6, gives coefficient of normalized cross correlation,

$$\gamma = \frac{\sum_{x,y} (f(x,y) - \bar{f}_{u,v})(t(x-u,y-v) - \bar{t})}{\sqrt{\sum_{x,y} (f(x,y) - \bar{f}_{u,v})^2} \sqrt{\sum_{x,y} (t(x-u,y-v) - \bar{t})^2}} \quad (6)$$

Whereas, the mean value of $f(x,y)$ is given as below by Eq. 7.

$$\bar{f}_{x,y} = \frac{1}{N_x N_y} \sum_{x=u}^{u-1+N_x} \sum_{y=v}^{v-1+N_y} f(x,y) \quad (7)$$

Same data was used with cross correlation, the image recognition from same signer is done more efficiently as compared to simple normalized correlation. However as is evident from above equations, normalized cross correlation uses more calculations in terms of divisions, square roots etc.

5. Results and discussion

In this section, the performance of real time static hand gesture recognition system is evaluated for each of the 37 hand gesture 370 samples are stored in dataset. The input image is used for the performance evaluation. The suggested solution faces following problems:

Rotation: To solve problem of rotation invariance, the position of wrist is detected and taken as reference point. Fig. 8 shows detection of wrist as starting point for image recognition.

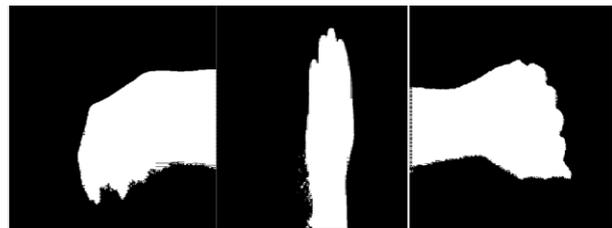


Fig. 8: Detection of wrist in different orientations

Scaling: The handling of scaling is left to next phase of this research. The images to be compared are brought to same size using the morphological operation of dilation.

Nonuniform illumination: Solution to nonuniform illumination is removal of back ground, the image we get after the subtraction of background is free from the problem of nonuniform illumination. Fig. 9 shows how an image changes once its background is removed.

Shadow: It is another problem with image recognition systems. For the time being it is ensured that images especially those in dataset are free for shadow. Moreover while segmenting the image skin detection technique is used, this reduces the effect of shadow in the image matching process.

6. Conclusion and future work

This paper is an attempt to solve the problem of image recognition in an unsupervised environment. The methodology used here is based on the concept of normalized correlation and normalized cross-correlation. The suggested solution exists for the situations where there are not more than one signers e.g. on a mobile device. The problem with the approach is that to make it scale invariant, rotation

invariant and background independent some extra measures are to be taken.

Future work involves improved solution for shadow removal and scaling invariance. Also, combining correlation with the entropy of the image is another area that can be used for efficiency.



Fig. 9: Background removal

Acknowledgment

For this research I am indebted to Mr. Nadeem Ali, lecturer Lahore Garrison University, and Dr. Adnan Khan, Associate Professor, Lahore Garrison University. Without their help I would not have completed this paper. I am also indebted to my Honourable Vice-Chancellor Major General (R) Obaid Bin Zakria, HI(M) for his policies to continuously motivate us for research.

References

Abdalla MS and Hemayed EE (2013). Dynamic hand gesture recognition of arabic sign language using hand motion trajectory features. *Global Journal of Computer Science and Technology*, 13(5-F): 27-33.

Abdo MZ, Hamdy AM, Salem SAER, and Saad EM (2015). Arabic alphabet and numbers sign language recognition. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 6(11): 209-214.

Albelwi NR and Alginahi YM (2012). Real-time arabic sign language (arsl) recognition. In the 5th International Conference on Communications and Information Technology, Cairo, Egypt: 497-501.

Ali SA (2013). Detection of urdu sign language using harr algorithms. *International Journal of Inventive Engineering and Sciences*, 1(6): 50-54.

Alvi AK, Azhar MYB, Usman M, Mumtaz S, Rafiq S, Rehman R, and Ahmed I (2007). Pakistan sign language recognition using statistical template matching. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 1(3): 765-768.

Bansal M, Saxena S, Desale D, and Jadhav D (2011). Dynamic gesture recognition using hidden markov model in static background. *International Journal of Computer Science Issues*, 8(6): 391-398.

Bhuyan MK, Kumar DA, MacDorman KF, and Iwahori Y (2014). A novel set of features for continuous hand gesture recognition. *Journal on Multimodal User Interfaces*, 8(4): 333-343.

Darwish SM, Madbouly MM, and Khorsheed MB (2016). Hand gesture recognition for sign language: A new higher order fuzzy HMM approach. *International Journal of Engineering and Technology*, 8(3): 157-164.

Fernandez JA, Boddeti VN, Rodriguez A, and Kumar BV (2015). Zero-aliasing correlation filters for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8): 1702-1715.

Jain C (2016). A novel technique for gesture recognition system. *International Journal of Recent Research Aspects*, 3(2): 21-24.

Jalilian B and Chalechale A (2014). Persian sign language recognition using radial distance and fourier transform. *International Journal of Image, Graphics and Signal Processing*, 6(1): 40-46.

Jie Z, Liang X, Feng J, Jin X, Lu W, and Yan S (2016). Tree-structured reinforcement learning for sequential object localization. In the 30th Conference on Neural Information Processing Systems, Barcelona, Spain: 127-135.

Kausar S, Javed MY, and Sohail S (2008). Recognition of gestures in pakistani sign language using fuzzy classifier. In the 8th WSEAS International Conference on Signal Processing, Computational Geometry and Artificial Vision, Rhodes, Greece: 101-105.

Moghaddam M, Nahvi M, and Hasanzadeh PR (2011). Static persian sign language recognition using kernel-based feature extraction. *International Journal of Information and Communication Technology Research*, 4(1): 21-28.

Mohandes M, Deriche M, and Liu J (2014). Image-based and sensor-based approaches to Arabic sign language recognition. *IEEE Transactions on Human-Machine Systems*, 44(4): 551-557.

Nachamai M (2013). Alphabet recognition of american sign language: A hand gesture recognition approach using sift algorithm. *International Journal of Artificial Intelligence and Applications*, 4(1): 105-115.

Nagi J, Nagi F, and Ahmed SK (2008). A MATLAB based Face recognition system using image processing and neural networks. In the 4th International Colloquium on Signal Processing and its Applications, Kuala Lumpur, Malaysia: 83-88.

Naoum R, Owaied H, and Joudeh S (2012). Development of a new arabic sign language recognition using k-nearest neighbor algorithm. *Journal of Emerging Trends in Computing and Information Sciences*, 3(8): 1173-1178.

Ohn-Bar E and Trivedi MM (2014). Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations. *IEEE Transactions on Intelligent Transportation Systems*, 15(6): 2368-2377.

Pandey P and Jain V (2015). Hand gesture recognition for sign language recognition: A review. *Hand*, 4(3): 4464-4470.

Pansare JR, Gawande SH, and Ingle M (2012). Real-time static hand gesture recognition for american sign language (ASL) in complex background. *Journal of Signal and Information Processing*, 3(03): 364-367.

Parveen D, Sanyal R, and Ansari A (2011). Clause boundary identification using classifier and clause markers in Urdu language. *Polibits Research Journal on Computer Science*, 43: 61-65.

Raes M, Ullah S, Rahman SU, and Rabbi I (2016). Image based recognition of Pakistan sign language. *Journal of Engineering Research*, 4(1): 21-41.

Sahoo AK, Mishra GS, and Ravulakollu KK (2014). Sign language recognition: State of the art. *ARPJ Journal of Engineering and Applied Sciences*, 9(2): 116-134.

Saqib S and Kazmi SAR (2017). Repository of static and dynamic signs. *International Journal of Advanced Computer Science and Applications*, 8(11): 101-105.

Sarkalehl AK, Poorahangaryan F, Zan B, and Karami A (2009). A neural network based system for persian sign language recognition. In the IEEE International Conference on Signal and Image Processing Applications, IEEE, Kuala Lumpur, Malaysia: 145-149. <https://doi.org/10.1109/ICSIPA.2009.5478627>

Sasirekha D and Chandra E (2012). Enhanced techniques for PDF image segmentation and text extraction. *International Journal of Computer Science and Information Security*, 10(9): 1-5.

- Shahzad N, Paulson B, and Hammond T (2009). Urdu Qaeda : Recognition system for isolated urdu characters. IUI 2009 Workshop on Sketch Recognition, Sanibel Island, Florida Chair: 1-5.
- Shurong L, Yuanyuan H, Zuojin H, and Qun D (2015). Key frame detection algorithm based on dynamic sign language video for the non specific population. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 8(12): 135-148.
- Singha J and Das K (2013). Indian sign language recognition using eigen value weighted euclidean distance based classification technique. *International Journal of Advanced Computer Science and Applications*, 4(2): 188-195.
- Singha J and Laskar RH (2017). Hand gesture recognition using two-level speed normalization, feature selection and classifier fusion. *Multimedia Systems*, 23(4): 499-514.
- Sykora P, Kamencay P, and Hudec R (2014). Comparison of SIFT and SURF methods for use on hand gesture recognition based on depth map. *AASRI Procedia*, 9: 19-24. <https://doi.org/10.1016/j.aasri.2014.09.005>
- Szegedy C, Toshev A, and Erhan D (2013). Deep neural networks for object detection. In the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, 2: 2553-2561.
- Tauseef H, Fahiem MA, and Farhan S (2009). Recognition and translation of hand gestures to urdu alphabets using a geometrical classification. In the 2nd International Conference in Visualisation, IEEE: 213-217. <https://doi.org/10.1109/VIZ.2009.11>
- Tehsin S, Masood A, Kausar S, and Javed Y (2013). Text localization and detection method for born-digital images. *IETE Journal of Research*, 59(4): 343-349.
- Zhang D, Maei H, Wang X, and Wang YF (2017). Deep reinforcement learning for visual object tracking in videos. Available online at: <http://arxiv.org/abs/1701.08936>
- Zhou B, Lapedriza A, Xiao J, Torralba A, and Oliva A (2014). Learning deep features for scene recognition using places database. In the 27th International Conference on Advances in Neural Information Processing Systems, MIT Press, 487-495.