A Real-Time Vision Based System for Recognition of Static Dactyls of Albanian Alphabet

Eriglen Gani
Department of Informatics
Faculty of Natural Sciences
University of Tirana
eriglen.gani@fshn.edu.al

Alda Kika
Department of Informatics
Faculty of Natural Sciences
University of Tirana
alda.kika@fshn.edu.al

Bruno Goxhi
Department of Informatics
Faculty of Natural Sciences
University of Tirana
bruno.goxhi@fshnstudent.info

Abstract

The aim of the paper is to present a real-time vision based system that is able to recognize static dactyls of Albanian alphabet. We use Kinect device, as an image receiving technology. It has simplified the process of vision based object recognition, especially for segmentation phase. Different from hardware based methods, our approach does not require that signers wear extra objects like data gloves. Two pre-processing techniques, including border extraction and image normalization have been applied in the segmented images. Fourier transform is applied in the resultant images which generates 15 Fourier coefficients representing uniquely that gesture. Classification is based on a similarity distance measure like Euclidian distance. Gesture with the lowest distance is considered as a match. Our system achieved an accuracy of 72.32% and is able to process 68 frames per second.

1 Introduction

Sign language is used as a natural way of communication between hearing impaired people. It is very important for the inclusion of deaf people in society. There exist a gap in communication between hearing impaired people and hearing ones. It comes from inability of hearing people to understand sign language. To overcome this gap most of the times interpreters can be used. The other, more comfortable solution is usage of technology. Natural interfaces can be used to capture the signs and understand their meaning by using body positions, hand trajectories and head movements. Using technology for catching, processing and translating dactyls in an understandable form for non deaf people, would help deaf ones integrate faster in the society [GK16]. A real time dactyls translator system would provide many facilities for this community. Many countries have tried to develop real-time sign language translator like: [Ull11, GK15c, TL11]. Unfortunately Albanian sign language (AlbSL) did not get much focus as other languages.

Deaf people in Albania used to communicate on the way that is based on finger-spelled Albanian words [ANA13]. Although not so efficient, dactyls play an important role in this type of communication. They form the bases of communication for deaf people. Albanian dactyl alphabet is composed of 36 dactyls. Among them 32 are static dactyls. 4 of them are dynamics ones, which are obtained from consecutive sequences of frames. The dynamic dactyls include (Ç, Ë, SH and ZH) [ANA13]. Our work is focused only in 32 static dactyls.

Two most widely used methods for building real-time translator system are hardware based and vision based [ZY14]. In hardware based method the signers have to wear data gloves or some other marker devices. It is not very natural to them. Vision

based methods are more challenging to be developed but are more natural for deaf people. Two most common problems include a)complex background and b)illumination change [ZY14]. Sometimes it is hard to distinguish human hands from other objects parts of the same environment. Sometimes the shadow or light effects the correct identification of human hand. Kinect sensor by Microsoft, has simplified the process of vision based object recognition, especially the segmentation phase. It offers some advantages like: provides color and depth data simultaneously, it is inexpensive, the body skeleton can be obtained easily and it is not effected by the light [GK16]. We are using Kinect sensor as a real-time image receiving technology for our work.

Our Albanian sign language translator system includes a limited set of number signs and dactyls. In the future other numbers, dynamic dactyls and signs will be integrated by making this system usable in many scenarios that require participation of deaf people. One usage of the system includes a program in a bar that could help the deaf people making some orders by combining numbers and dactyl gestures.

Till now there does not exist any gesture data set for Albanian sign language. We are trying to built a system that is able to translate static dactyls signs for Albanian sign language and in the future it will be extended to dynamic dactyls and other signs. Creating and continuously adding new signs to an Albanian gesture data set would help building a more reliable and useful recognition system for our sign language.

Section 1 gives a brief introduction. Section 2 summarizes some related works. The rest of the paper is organized as follows. Section 3 presents an overview of methodology and a brief description of each methodology's processes. Section 4 describes the experimental environment. Section 5 presents the experiments and results. The paper is concluded in Section 6 by presenting the conclusions and future work.

2 Related Work

Many researchers have followed different methodologies for building sign language recognition systems. They are categorized into several types based on input data and hardware dependency. Signs, which are mostly performed by human hands can be static or dynamic. The sign language recognition systems are categorized as hardware based or vision based.

Many works have been done to integrate some hardware based technologies to capture and translate sign gestures, among them the most widely used are data gloves. Authors at [Sud14] built a portable system for deaf people using a smart glove capable of capturing finger movements and hand movements. In general data gloves achieve high performance but are expensive and not a proper way to human-computer interaction perspective [GK15b].

Web cameras with an image processing system can be used in vision based approaches. Research at [SSKK16] presents a vision based methodology using web cameras to recognize gesture from Indian sign language. The system achieves high recognition rate. Authors at [WKSE02] and [LGS08] use color camera to capture input gestures and then SVM (Support Vector Machine) and Fuzzy C-Means respectively to classify hand gestures. Despite this, in general web cameras generate low quality of images and have an inability to capture other body parts. It is also hard to generalize the algorithms for web cameras due to many different shapes and colors of hands [GK15b].

Kinect sensors by Microsoft has simplified the process of vision based object recognition. It has many advantages as: provide color and depth data simultaneously, it is inexpensive, the body skeleton can be obtained easily and it is not effected by the light. Various researchers are using Microsoft Kinect sensor for sign language recognition as in [GK15c], [SB13], [VAC13].

Vision based hand gesture recognition provides more intuitive interaction with a system. It is a challange task to identify and classify hand gesture. Shape and movement play an important role in gesture categorization. A comparison between two most widely used algorithm for shape recognition is done at [CBM07]. It compares Fourier descriptors (FD) and HU moments in terms of performance and accuracy. Algorithms are compared against a custom and a real-life gesture vocabulary. Experiment results show that FD is more efficient in terms of accuracy and performance.

Research at [BGRS11] addresses the issue of feature extraction for gesture recognition. It compares Moment In-variants and Fourier descriptors in terms of in-variance to certain transformations and discrimination power. ASL images were used to form gesture dictionary. Both approaches found difficult to classify correctly some classes of ASL.

Authors at [BF12] compare different methods for shape representation in terms of accuracy and real-time performance. Methods that were used to compare them include region based moments (Hu moments and Zenike moments) and Fourier descriptors. Conclusions showed that Fourier descriptors have the highest recognition rate.

Shape is an important factor for gesture recognition. There exist many methods for shape representation and retrival. Among them Fourier descriptors achieve good representation and normalization. Authors at [ZL$^+$02] compare different shape signatures used to derived Fourier descriptors. Among them: complex coor-

dinates, centroid distance, curvature signature and cumulative angular function. Article concludes that centroid distance is significantly better than other three signatures.

Sign language is not limited only in static gesture. Majority of signs are dynamic ones. Research at [RMP+15] proposed a hand gesture recognition method using Microsoft Kinect. It uses two different classification algorithms DTW and HMM by discussing the pros and cons of each technique.

3 Methodology

Figure 1 gives an overview of the followed methodology for our work.

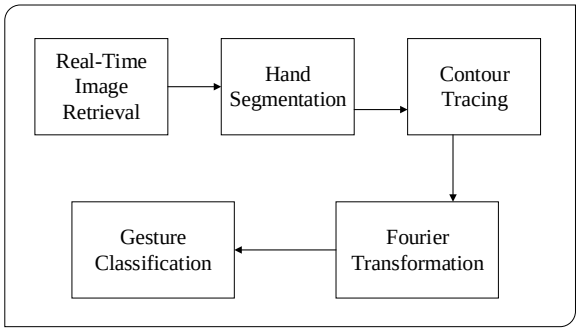

Figure 1: Methodology

Microsoft Kinect is used as a real-time image retrieval. Kinect consists of an RGB camera, an IR emitter, an IR depth sensor, a microphone array and a tilt. The RGB camera can capture three-channel data in a 1280 x 960 resolution at 12 FPS or a 640 x 480 resolution at 30 FPS. In our work images consist of a 640 x 480 resolution at 30 FPS. The valid operating distance of Kinect is approximately 0.8m to 4m [MSD16]. Due to its advantages, it has simplified the process of vision based object recognition, especially for segmentation phase.

Every pixel generated from Kinect device contains information of their depth location layer and player index. By using player index we focus only in pixels that are part of human body [WA12]. In this way all other pixels, not part of human body are excluded. By applying a constant threshold we can obtain the human hand, since it is the first part of human body towards the Kinect device [GK15a].

In order to perform Fourier transform we have to generate a centroid function which is based in hand image contour. Theo Pavlidis is used as a hand contour tracking algorithm [Pav12]. The segmented hand is transformed in greyscale where each pixel is classified as a white or a black one. After applying Theo Pavlidis algorithm, the resultant image contains only border pixels of human hand.

Fourier descriptors can be derived from complex coordinates, centroid distance, curvature signature or cumulative angular function. In our case centroid distance is used due to [ZL+02]. After locating the center of white pixels in the image, we have calculated the distance of every border pixels from it. It gives the centroid function which represents two dimensions area.

The normalization process consists of extracting the same number of pixel, equally distributed, among hand border. Choosing a lower number of border pixels decrease the system accuracy, while choosing a higher number decrease the system performance. In our case a number of 128 pixels has been chosen.

Fourier descriptors are used to transform the resultant image into a frequency domain. For each image, only the first 15 Fourier coefficients are used to define them uniquely. Other Fourier coefficients do not effect system accuracy. Every input gesture is compared against a training data set using a similarity distance measure like Euclidian distance. The gesture with the lowest distance is considered as a match.

4 Experiment Environment

Experiment environment used for implementing and testing our real-time static dactyls recognition system is composed of the following hardware: A notebook with a processing capacity of 2.5 GHz, Intel Core-i5. A memory capacity of 6 GB of RAM and a Windows 10 operating system with a 64-bit architecture. Microsoft Kinect for Xbox 360 is used as a real-time image retrieval technology. It generates 30 frames per second and can be used as a RGB camera and also can provide depth data.

System was developed using .Net technology. Kinect for Windows SDK 1.8.0.0 was used as library between Kinect device and our application. It provides a way to process Kinect signals. An overview of the system architecture is given at Figure 2.

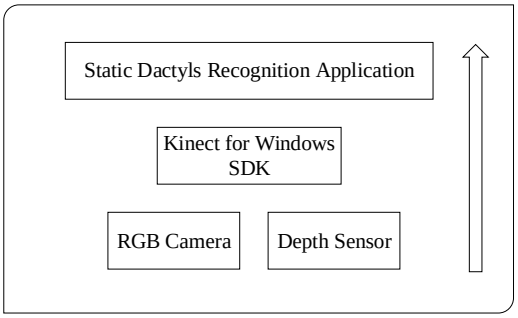

Figure 2: System Architecture

5 Experiment and Results

To test the proposed system, several experiments were conducted. Each experiment is based on two aspects: accuracy and computation latency. The first experiment measured the accuracy of correct identification and classification of static dactyls. Our system is not able to identify and classify dynamic dactyls. It is based only in static ones. Firstly training data set is created. It contains 320 dactyl gestures taken from two different signers. Each gesture is performed 5 times from each signers and is represented by 15 Fourier coefficients. There are in total 4800 coefficients (15x320). For real-time testing, 4 different signers were used. Each signer performed 5 gestures for each dactyl sign. In total they performed 640 experiments. Each element in testing data set is compared against all elements in training data set. The element with lowest Euclidian distance is considered as a match. The average recognition accuracy for each static dactyl is given in the Table 1 and Table 2.

Table 1: Average Recognition Accuracy of Testing Data Set

Static Dactyl	True Recognition Rate (%)	False Recognition Rate (%)
A	100	0
B	87.5	12.5
C	70	30
D	67	33
DH	72.5	27.5
E	70	30
F	60	40
G	75	25
GJ	72	28
H	87.5	12.5
I	82.5	17.5
J	98.75	1.25
K	100	0
L	50	50
LL	52.5	47.5
M	85	15
N	55	45

For all static dactyls, the system achieves an average accuracy rate of 72.32%. Results show that dactyls with the highest accuracy rate are 'A', 'J', 'K' and 'V'. Their accuracy rate is above 95%. Dactyls with the lowest accuracy rate are 'L', 'Y' and 'Z'. Their accuracy rate is below 52%.

Table 3 and Table 4 give information regarding static dactyls confusion percentages. Some of Albanian dactyls are easily confused with other dactyls due to their similarity. Based on experimental results

Table 2: Average Recognition Accuracy of Testing Data Set

Static Dactyl	True Recognition Rate (%)	False Recognition Rate (%)
NJ	87.5	12.5
O	65	35
P	54	46
Q	52.5	47.5
R	65	35
RR	85	15
S	84	16
T	67.5	32.5
TH	54	46
U	67.5	32.5
V	97.5	2.5
X	87.5	12.5
XH	60	40
Y	50	50
Z	52	48

dactyls "D", "E", "F", "N", "O" are more confused ones.

The second experiment deals with system performance. We want to achieve a performance that allows the system to be deployed in real-time. For every sign we analyzed the time required for the following phases: hand segmentation, hand contour tracing, normalization, centroid function generation, Fourier transformation and gesture classification. Table 5 summarize the results.

The system needs approximately 12 to 17 ms to process a static dactyl. Most of the overall time is consumed by hand segmentation and gesture classification processes. They occupy approximately 82% of total time. It can be deployed without any latency in a real-time system that uses Microsoft Kinect.

6 Conclusion and Future Work

The aim of this paper is to built a real-time system that is able to recognize static dactyls for Albanian alphabet by using Microsoft Kinect. Albanian alphabet is composed of 36 dactyls and 32 of them are static. The static dactyls are used as inputs for our system. Kinect device provides a vision based approach and is used as an image retrieval technology. Its main feature includes depth sensor. For every static dactyl, a data set with 15 Fourier coefficients was built. In total data set consists of 4800 coefficients. For testing purpose, 4 different signers were used. Each of them performed 5 times each of the static dactyls. A total of 640 experiments were conducted. For classification purpose a similarity distance measures like Euclidian

Table 3: Confusion Dactyls Percentages (%)

Static Dactyl	Confusion Dactyls Percentages
A	{A,100}
B	{B,87.5}; {TH,12.5}
C	{C,70}; {E,12}; {X,12}; {Y,6};
D	{B,20}; {D,67}; {DH,5}; {H,5}; {U,3};
DH	{DH,72.5}; {E,2.5}; {F,12.5}; {U,12.5};
E	{C,2.5}; {E,70}; {J,2.5}; {TH,12.5}; {V,12.5}
F	{B,12.5}; {DH,12.5}; {F,60}; {TH,2.5}; {U,12.5}
G	{G,75}; {I,12.5}; {J,12.5}
GJ	{GJ,72}; {X,8}; {Z,20}
H	{B,5}; {F,5}; {H,87.5}; {TH,2.5};
I	{I,82.5}; {J,15}; {Y,2.5}
J	{I,1.25}; {J,98.75}
K	{K,100}
L	{L,50}; {XH,35}; {Z,15}

Table 4: Confusion Dactyls Percentages (%)

Static Dactyl	Confusion Dactyls Percentages
LL	{GJ,32.5}; {LL,52.5}; {M,15}
M	{E,12.5}; {M,85}; {Y,2.5}
N	{A,12.5}; {N,55}; {O,15}; {Q,15}; {Z,2.5};
NJ	{GJ,12.5}; {NJ,87.5}
O	{C,2.5}; {N,2.5}; {O,65}; {P,12.5}; {Q,17.5};
P	{P,54}; {O,23}; {R,23}
Q	{N,25}; {O,22.5}; {Q,52.5}
R	{J,18}; {N,17}; {R,65}
RR	{J,15}; {RR,85}
S	{S,84}; {Z;16}
T	{P,32.5}; {T,67.5}
TH	{DH,15}; {P,16}; {Q,15}; {TH,54}
U	{N,32.5}; {U,67.5}
V	{E,2.5}; {V,97.5}
X	{K,12.5}; {X,87.5}
XH	{DH,12.5}; {TH,15}; {X,12.5}; {XH,60}
Y	{A,25}; {J,25}; {Y,50}
Z	{DH,20}; {F,14}; {N,14}; {Z,52}

Table 5: Computational Latency Results

Phases	Computation Latency (ms)
Hand Segmentation	5.49308
Hand Contour Tracing	0.08145
Normalization	0.03261
Centroid Function Generation	0.03705
Fourier Transformation	2.45306
Gesture Cassification	6.67130
Total	**14.7686**

distance was used. Every element in testing data set is compared against each element in training data set. The element with the lowest Euclidian distance is considered as a match. The system is tested against accuracy and performance. Based on experiments results the system achieves an accuracy rate of 72.32%. The system needs to compute a static dactyl is 14.05 ms in average. It can be deployed in a image receiving technology that generates 68 frames per second.

Future work consists of improving the overall system performance and accuracy by applying a more reliable data set. This can be done by including more diverse signers who have high knowledge of Albanian sign language. The future work also consist of adding dynamic dactyls as well as other gestures of Albanian sign language.

References

[ANA13] ANAD. *Gjuha e Shenjave Shqipe 1.* Shoqata Kombëtare Shiptare e Njerëzve që nuk Dëgjojnë, 2013.

[BF12] Salah Bourennane and Caroline Fossati. Comparison of shape descriptors for hand posture recognition in video. *Signal, Image and Video Processing*, 6(1):147–157, 2012.

[BGRS11] Andre LC Barczak, Andrew Gilman, Napoleon H Reyes, and Teo Susnjak. Analysis of feature invariance and discrimination for hand images: Fourier descriptors versus moment invariants. In *International Conference Image and Vision Computing New Zealand IVCNZ2011*, 2011.

[CBM07] Simon Conseil, Salah Bourennane, and Lionel Martin. Comparison of fourier descriptors and hu moments for hand posture recognition. In *Signal Processing Conference, 2007 15th European*, pages 1960–1964. IEEE, 2007.

[GK15a] Eriglen Gani and Alda Kika. Identifikimi i dores nepermjet teknologjise microsoft kinect. *Buletini i Shkencave te Natyres*, 20:82–90, 2015.

[GK15b] Eriglen Gani and Alda Kika. Review on natural interfaces technologies for designing albanian sign language recognition system. *The Third International Conference On: Research and Education Challenges Towards the Future*, 2015.

[GK15c] Archana S Ghotkar and Gajanan K Kharate. Dynamic hand gesture recognition and novel sentence interpretation algorithm for indian sign language using microsoft kinect sensor. *Journal of Pattern Recognition Research*, 1:24–38, 2015.

[GK16] Eriglen Gani and Alda Kika. Albanian sign language (AlbSL) number recognition from both hand's gestures acquired by kinect sensors. *International Journal of Advanced Computer Science and Applications*, 7(7), 2016.

[LGS08] Yun Liu, Zhijie Gan, and Yu Sun. Static hand gesture recognition and its application based on support vector machines. In *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2008. SNPD'08. Ninth ACIS International Conference on*, pages 517–521. IEEE, 2008.

[MSD16] MSDN. Kinect for windows sensor components and specifications, April 2016.

[Pav12] Theodosios Pavlidis. *Algorithms for graphics and image processing*. Springer Science & Business Media, 2012.

[RMP+15] J.L. Raheja, M. Minhas, D. Prashanth, T. Shah, and A. Chaudhary. Robust gesture recognition using kinect: A comparison between DTW and HMM. *Optik - International Journal for Light and Electron Optics*, 126(11-12):1098–1104, jun 2015.

[SB13] Kalin Stefanov and Jonas Beskow. A kinect corpus of swedish sign language signs. In *Proceedings of the 2013 Workshop on Multimodal Corpora: Beyond Audio and Video*, 2013.

[SSKK16] S Shruthi, KC Sona, and S Kiran Kumar. Classification on hand gesture recognition and translation from real time video using svm-knn. *International Journal of Applied Engineering Research*, 11(8):5414–5418, 2016.

[Sud14] Bh Sudantha. A portable tool for deaf and hearing impaired people, 2014.

[TL11] Pedro Trindade and Jorge Lobo. Distributed accelerometers for gesture recognition and visualization. In *Technological Innovation for Sustainability*, pages 215–223. Springer, 2011.

[Ull11] Fahad Ullah. American sign language recognition system for hearing impaired people using cartesian genetic programming. In *Automation, Robotics and Applications (ICARA), 2011 5th International Conference on*, pages 96–99. IEEE, 2011.

[VAC13] Harsh Vardhan Verma, Eshan Aggarwal, and Swarup Chandra. Gesture recognition using kinect for sign language translation. In *Image Information Processing (ICIIP), 2013 IEEE Second International Conference on*, pages 96–100. IEEE, 2013.

[WA12] Jarrett Webb and James Ashley. *Beginning Kinect Programming with the Microsoft Kinect SDK*. Apress, 2012.

[WKSE02] Juan Wachs, Uri Kartoun, Helman Stern, and Yael Edan. Real-time hand gesture telerobotic system using fuzzy c-means clustering. In *Automation Congress, 2002 Proceedings of the 5th Biannual World*, volume 13, pages 403–409. IEEE, 2002.

[ZL+02] Dengsheng Zhang, Guojun Lu, et al. A comparative study of fourier descriptors for shape representation and retrieval. In *Proc. 5th Asian Conference on Computer Vision*. Citeseer, 2002.

[ZY14] Yanmin Zhu and Bo Yuan. Real-time hand gesture recognition with kinect for playing racing video games. In *2014 International Joint Conference on Neural Networks (IJCNN)*. Institute of Electrical & Electronics Engineers (IEEE), jul 2014.