Res. Lett. Inf. Math. Sci., 2005, Vol. **7**, pp 127-134
Available online at http://iims.massey.ac.nz/research/letters/

A Color Hand Gesture Database for Evaluating and Improving Algorithms on Hand Gesture and Posture Recognition

FARHAD DADGOSTAR, ANDRE L. C. BARCZAK, ABDOLHOSSEIN SARRAFZADEH

Institute of Information & Mathematical Sciences
Massey University at Albany, Auckland, New Zealand [1]

Abstract: With the increase of research activities in vision-based hand posture and gesture recognition, new methods and algorithms are being developed. Although less attention is being paid to developing a standard platform for this purpose. Developing a database of hand gesture images is a necessary first step for standardizing the research on hand gesture recognition. For this purpose, we have developed an image database of hand posture and gesture images. The database contains hand images in different lighting conditions and collected using a digital camera. Details of the automatic segmentation and clipping of the hands are also discussed in this paper.

Key Words: Hand Posture Detection, Hand Image Database, Gesture Recognition, Hand Posture Recognition.

1. Introduction

Detecting and understanding hand and body gestures is becoming a very important and challenging task in computer vision. The significance of the problem can be easily illustrated by the use of natural gestures that we apply together with verbal and nonverbal communications. The use of hand gestures in support of verbal communication is so deep that they are even used in communications that people have no visual contact (e.g. talking on the phone).

There are different approaches for recognizing hand gestures in computer science community, some of which require wearing marked gloves or attaching extra hardware to the body of the subject. These approaches are from the users' point of view considered to be intrusive, and therefore less likely to apply for real world applications. On the other hand, vision-based approaches considered non-intrusive and therefore more likely to be used for real world applications.

A prerequisite of a vision-based gesture and hand posture recognition is the detection of the hand in the image. In addition, recognition of the shape of the hand is also important in some applications like Sign Language recognition and Human Computer Interaction. The 2D representation of the hand in vision-based systems, as an articulated object makes its geometry varying over time, and consequently makes the problem harder to solve in comparison to similar problems (e.g. face detection).

[1] F.Dadgostar@Massey.ac.nz , A.L.Barczak@Massey.ac.nz , A.H.Sarrafzadeh@Massey.ac.nz

In this area of research many problems still need to be solved. Robust and fast recognition of the hand over time and space are probably the most important problems. Unfortunately the research done on vision-based gesture recognition is very limited and there is no publicly available general purpose benchmarking data or image database for gesture recognition. This paper discusses a hand gesture database that we have developed for this purpose. In section 2, discusses some specific purpose databases that are available on the web, and section 3 describes the Massey Hand Gesture Database that has been developed by the authors of this article. In section 4 we describe one of the applications of this database, and sections 5 and 6 describe the possible features that need to be added to the hand gesture database, and the results of the work, respectively.

2. Existing Databases

According to the literature, and best of our knowledge, there are a few publicly available gesture image databases. Athitsos and Sclaroff (Athitsos & Sclaroff, 2003) published a database for hands posed in different gestures. The database contains more than 107000 images. Despite the fact that the database covers 26 gestures and has ground truth tables, the images actually present only the edges of the hands. Tests for algorithms that are not base on edges are not feasible. Athitsos also contributed to the creation of an American Sign Language (ASL) video sequence database (Cited in Table 1). These videos present the upper body part of a person signaling short texts in ASL. There are images from 4 different cameras. The videos were recorded in a rate of 60 frames per second. Some frames present the hands in a small scale and they are sometimes blurred. It is also difficult to cluster sets of hands where the gesture is of a certain type. This database would be suitable for testing detection algorithms, but it would be difficult to use those images for training. There are also some other databases that are not specifically related to gesture but are particularly related to the subject. The summary of some of the databases that were cited in the literature indicated in Table 1.

Table 1. Image databases were addressed in the literature

Title	Address	Research Institute	Description
color face image database	http://dsp.ucd.ie/~prag/	University of Dublin	A collection images containing face gathered from different sources
AT&T Laboratories Cambridge face database	http://www.uk.research.att.com/facedatabase.html	AT&T	10 different images from 40 different subjects
Video sequences of American Sign Language (ASL)	http://csr.bu.edu/asl/html/sequences.html	Boston University	
Compaq image database	This database was addressed in several research, but is not publicly available anymore	Compaq Inc.	A collection of images and ground truth data of the skin segments

2.1. The need for a new database on hand gestures

As new hand detection and gesture recognition algorithms are being developed, the use of features such as color, size, and shape of the favorite object are more likely to be used. Currently available databases are either for special purposes, or suffer from the lack of the desired features (e.g. not having color or very small size of the samples).

Research shows that color is one of the important features in body tracking. Color can be found to be invariant to changes in size, orientation and sometimes occlusion. In addition, according to Moore's law: every 18 month the processing speed and available memory size doubles, and then it is possible that in the near future, using samples with higher details would be preferred by researchers.

3. The Massey Gesture Database

The Massey Hand Gesture Database is an image database containing a number of hand gesture and hand posture images. The database has been developed by the authors to evaluate their methods and algorithms for real-time gesture and posture recognition. It is posted on the web with the hope of assisting other researchers investigating in the related domains, and is available form the following web address:

http://www.massey.ac.nz/~fdadgost/xview.php?page=hand_image_database/default

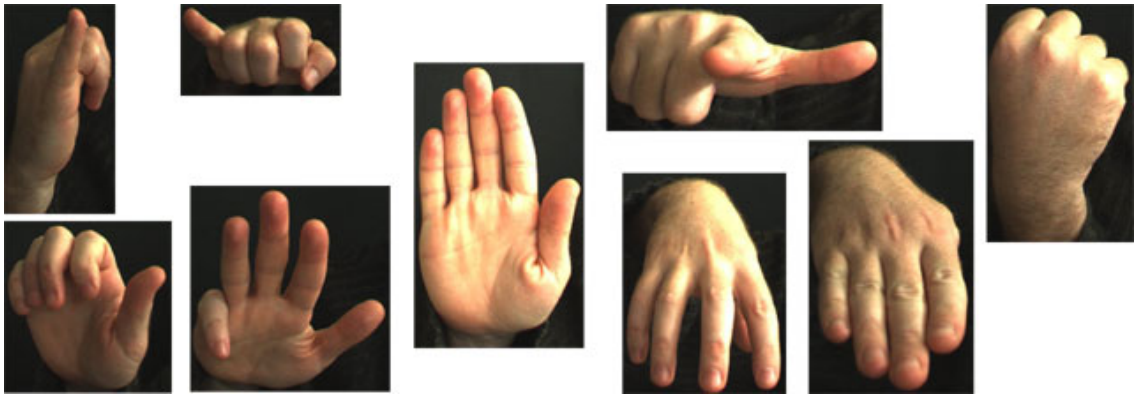

Figure 1: Some of the samples from the *Massey Hand Gesture Database*

At this stage, the Database includes about 1500 images of different hand postures, in different lighting conditions. The data was collected by a digital camera mounted on a tripod from a hand gesture in front of a dark background, and in different lighting environments, including normal light and dark room with artificial light. Together with the original images there is a clipped version of each set of images that contains only the hand image. The maximum resolution of the images is 640x480 with 24 bit RGB color (Figure 1).

So far, the database contains material gathered from 5 different individuals. Currently 6 sets of data in JPEG format are available publicly, as follows:

Table 2: List of datasets

	Dataset	**Lighting Condition**	**Background**	**Size**	**Number of Images**
1	Hand gesture	Normal	Dark background	640x480	169
2	Hand gesture	Normal	RGB(0,0,0)	Varying/ Clipped	169
3	Hand palm	Normal	Dark background	640x480	145
4	Hand palm	Normal	RGB(0,0,0)	Varying/ Clipped	145

5	Hand palm	Artificial light/ Dark room	Dark background	640x480	498
6	Hand palm	Artificial light/ Dark room	Dark background	Varying/ Clipped	498

3.1. Background Elimination and Clipping Process

The hand images in two of the datasets have been segmented from the background. The background elimination process was done by an in-house developed application. This application is based on global hue filter that has been described in (Dadgostar, Sarrafzadeh, & Johnson, 2005; Dadgostar, Sarrafzadeh, & Overmyer, 2005). The global skin detector could segment perfectly, the hand from the background in about 95% of the images. For the rest of the images manual adjustments were done to produce better results. For those images that have no background, those pixels that have a value not equal to RGB(0,0,0) belong to the hand, so the rest of the pixels can be ignored. Therefore researchers can use this feature for further processing, or they can add their own background images to the dataset (Figure 2).

Figure 2: Background elimination

For clipping the hand from the rest of the image, a simple boundary detection algorithm was used to make the process faster (Figure 3). However, in some images the boundary detection algorithm didn't work properly (Figure 4), in this case the rectangle that indicates the hand was manually clipped. In overall, no other manipulation like shrinking or color balancing was applied to the images.

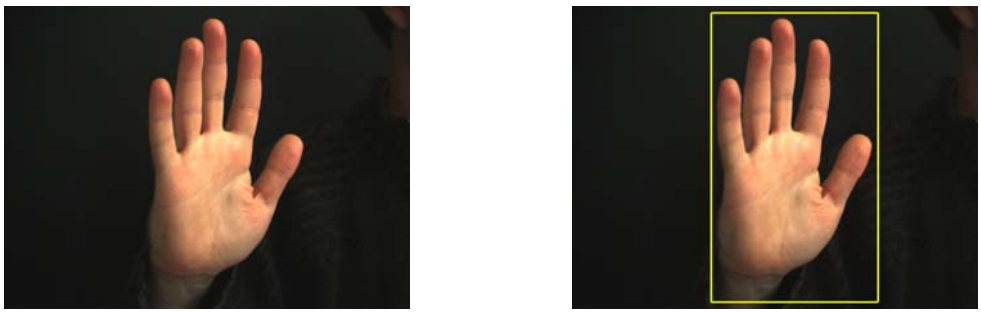

Figure 3: Hand's boundary detection

Figure 4: Cloth on the right side of the image makes automatic clipping difficult

3.2. Adding Random Backgrounds

In many experiments the same object will have to be analysed using images containing the same object with different backgrounds. For this reason it was decided to eliminate the background. Thus, it will be possible to add random backgrounds to the object, to enrich the original samples. Figure 5 shows three samples where a choice of random backgrounds were added to the images.

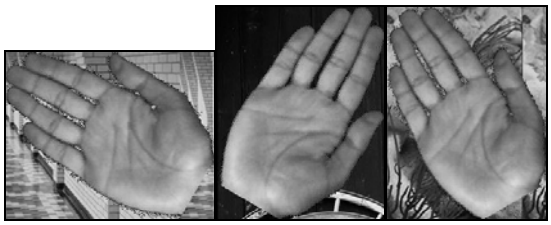

Figure 5: Adding random backgrounds

4. Database Application

One of the typical applications for an image database is to use it as a training set for learning algorithms. The same database could also be used for the testing phase, but it is more convenient to do tests with real images acquired separately from a different camera. Although this idea might be against the accepted procedures for testing classifiers, it gives a more conservative estimate of the performance. The patterns produced by different cameras in different environments may be quite different from the training set, so testing with these new images gives a clear picture of what to expect when using the classifier in the field. If the performance is not acceptable due to the background then more examples should be added to the training set.

The OpenCV implementation of Viola and Jones (Viola & Jones, 2001) method, extended by Lienhart et al. (Lienthart & Maydt, 2002), was used to train hands presented with a simple gesture. Due to the non-invariant nature of the features used by this method, articulation and rotation can cause problems for the classifier. Theoretically it is possible to train separate binary classifiers for each articulation and for each rotation. In practice this would be difficult to implement, as there are too many possibilities. However clustering gestures and rotations in smaller groups is feasible. It was decided to concentrate on rotation and using a single gesture for the preliminary tests.

Preliminary results showed that the tolerance to rotation is more critical for hand detection than it is for face detection. While faces can be rotated up to ±15 degrees, hands have to be within a margin of ±4 degrees to be successfully tracked by this algorithm. Kolsch and Turk (Kolsch & Turk, 2004) did experiments which showed the same tolerance for hand detection. The training experiments required 31 classifiers, one classifier for each angle. The interval was 6 degrees, within the tolerance required. A total of 145 base hand images with about 4400 images were used to train each cascade.
The experiments showed that the classifiers created for different angles, although created with the same twisted image samples, behaved very differently. The best hit ratio (92%) was for the classifier detecting hands in 30 degrees of rotation. The worst hit ratio was 37% for the classifier at -90 degrees. The false positive ratio was evenly spread. The classifiers around 0 up to 60 degrees were the best ones, while classifiers closer to 90 degrees had a poor performance. This can be explained as caused by differences in lighting conditions. The test images, actually created at angles close to 90 degrees, presented a different pattern from the ones used to train the classifiers. In future extensions of the database this factor has to be carefully considered. Although the original method does not use colours for gathering feature values, colours could be used in the future to speed up the searching process by eliminating regions of the image where the hands are certainty absent. A complete technical report can be seen in (Barczak & Dadgostar, 2005).

5. Extending the Database

The database could be extended in different ways. Extending the database could be an infinite procedure, but considering the author's research priorities, we have considered the following extensions to the database.
The gesture database itself, can be extended in three ways. Firstly the lighting conditions could be enriched by producing more of the same type of images. Secondly more gesture types could be collected. Finally, a larger number of different hands (from different people) would extend the database with respect to the geometrical proportions of the human hand.
The lighting conditions essentially force the algorithms to learn about the various patterns created in 2D image by a 3D object. Unless some special gadgets are used to control the lighting, it is very difficult to vary the positions of the light fairly along the three axis. For the first version of this database a simple lamp going around the subjects was used, but this is far from ideal.
The gestures set is virtually infinite, therefore it is difficult to choose the relevant ones. A possible approach to this problem is to constrain the set to some standard or known gestures, such as the American Sign Language. The existing databases for ASL are not necessarily suitable for training due to the lack of variations in lighting and subjects as well as their sizes.
Some algorithms would consider a long hand different than a short one. In order to generalize the database, many examples are needed. In addition, people tend to gesticulate in slightly different ways.
The problem of keeping the subjects in a still position in order to take a shot of the desired gesture in the correct lighting is a challenging one. Many images are discarded during the process because they are blurred or because the image was twisted in a

certain angle which is not easy to recover. One alternative to collecting data for training is to use 3D modelling tools. For face recognition this approach was very successful (Kouzani, 2003).

For hand recognition, a 3D modelling tool like Povray can be used for 3D hand modelling. Povray is a tool that is publicly available and it is easy to program (www.povray.org). This approach would have the advantage of automatically varying the lighting fairly in all directions and even produce very complex patterns of lighting by introducing more than one source of light. However some variations of the human hand shape (proportion, articulation limits etc) would have to be carefully considered to make the images as close as possible to reality. Figure 6 shows an example of a 3D model created exclusively using deformed cylinders and spheres. This image is far from ideal, but it illustrates that once the modelling is finished then 2D images can be generated in a much faster way than the slow process of photographing real hands.

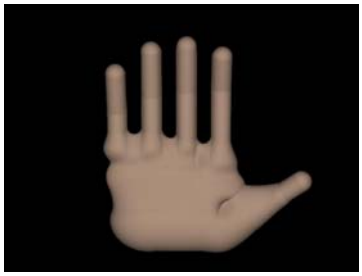

Figure 6: Basic hand image, generated using Povray

6. Conclusion

In this paper, we described the development of a color hand gesture database. The database has been developed with the intention of providing a common benchmark for the algorithms that are developed by the authors. It contains color images of the hand in different lighting conditions in front of a dark background. The purpose of using the dark background was the possibility to eliminate the background and to automatically segment the hand region. This feature enables researchers to add their own backgrounds to the image or to use it as an object with known boundaries.

References

Athitsos, V., & Sclaroff, S. (2003). *Estimating 3D Hand Pose from a Cluttered Image.* Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, USA.

Barczak, A. L. C., & Dadgostar, F. (2005). *Real-time Hand Tracking Using the Viola and Jones Method.* Paper presented at the IEEE International Conference on Image Processing (IASTED), Toronto, Canada.

Dadgostar, F., Sarrafzadeh, A., & Johnson, M. J. (2005). *An Adaptive Skin Detector for Video Sequences Based on Optical Flow Motion Features.* Paper presented at the International Conference on Signal and Image Processing (SIP), Hawaii, USA.

Dadgostar, F., Sarrafzadeh, A., & Overmyer, S. P. (2005). *An Adaptive Real-time Skin Detector for Video Sequences.* Paper presented at the International Conference on Computer Vision, Las Vegas, USA.

Kolsch, M., & Turk, M. (2004). *Analysis of Rotational Robustness of Hand Detection with a Viola_Jones Detector*. Paper presented at the International Conference on Pattern Recognition, Cambridge, U.K.

Kouzani, A. Z. (2003). Locating human faces within images *Comput. Vis. Image Underst. , 91* (3), 247-279

Lienthart, R., & Maydt, J. (2002). *An Extended Set of Harr-like Features for Rapid Object Detection.* Paper presented at the IEEE ICIP2002.
Viola, P., & Jones, M. J. (2001). *Robust Real-time Object Detection*: Cambridge Research Laboratory.