Transformation Invariance in Hand Shape Recognition

Thomas Coogan[1], Alistair Sutherland[2]
School of Computing, Dublin City University, Ireland
[1]*tcoogan@computing.dcu.ie* [2]*alistair@computing.dcu.ie*

Abstract

In hand shape recognition, transformation invariance is key for successful recognition. We propose a system that is invariant to small scale, translation and shape variations. This is achieved by using a-priori knowledge to create a transformation subspace for each hand shape. Transformation subspaces are created by performing Principal Component Analysis (PCA) on images produced using computer animation. A method to increase the efficiency of the system is outlined. This is achieved using a technique of grouping subspaces based on their origin and then organising them into a hierarchical decision tree. We compare the accuracy of this technique with that of the Tangent Distance technique and display the results.

1. Introduction

There have been many approaches to hand shape recognition [3,4,5,6]. All these recognition systems display sub-optimal results due to the highly deformable nature of the hand. Any hand shape recognition system needs to be able to cope with slight distortions along the 28 degrees of freedom of the hand.

Another significant problem of hand shape recognition is accurate segmentation of the hand from images. With these issues in mind one important question is how to align these object images. A commonly used approach is to align hand objects based on the centroid of the objects bounding box. However with the aforementioned variances of the hand in mind, this centroid is inconsistent. Therefore the system should also be capable of handling translated object images.

Simard et al. [9] proposed the tangent distance technique to achieve transformation invariance. This method approximates the high dimensional transformation hyper-plane with its tangent plane. Now two images can be compared by finding the distance between their transformation tangent planes. In section 3 we compare results from our subspace method with that of the Tangent Distance technique.

We propose a transformation subspace technique to combat these issues. Multiple subspace approaches have been employed previously by Wu [1], where the training data was produced in an ad-hoc manner, then sectioned into subspaces using a-priori knowledge. This method proposed an exhaustive search along subspaces for a given test image. Zhao [2] used an approach to calculate transformation subspaces from original subspaces created from 'perfect' training images for face recognition. They offer a multi-resolution search to speed up the exhaustive search of the test image to the original subspaces, along with each transformation subspace. It is important to note that this method only allows for 2D images transformations.

Our proposed method creates the invariant subspace from a sampled subset of all possible transformation images. These images are produced systematically using the commercially available POSER modelling software [10] and can include 3D hand transformation. Performing PCA on these images will generate a subspace that accurately represents the complex transformation hyper-plane. Instead of performing an exhaustive search on each subspace, we propose a hierarchical tree search that groups similar eigenspaces together. This is achieved by performing a fuzzy k-means algorithm on the origins of the eigenspaces. This allows us to reduce the search involved while retaining accurate search results.

The remainder of the paper is organised as follows: The method of obtaining the complete training set is reported in section 2. In section 3 we outline the proposed technique. Section 4 details some experimental results and comparisons of the transformation subspace distance with the tangent distance. We finish with the conclusion in Section 5.

2. Creating training database

One crucial factor in this system is how competently the training images can be produced to accurately represent the transformation hyper-plane.

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

Using computer animation we can accurately create a model of the hand in any orientation. It is also possible to change all of the characteristics of the hand, for example hand size, direction, orientation and skin colour along with lighting and scene conditions.

In this paper we concentrate on translation, rotation and small random hand configuration transformations. Once the original pose for each hand shape is manually initialised, all subsequent transformations can be generated automatically by manipulating the hand model using the PYTHON scripting tool provided by POSER.

Capturing an image of the hand model at each stance establishes the complete set of hand shape transformations needed to construct its transformation subspace. Using this method of a-priori knowledge to construct the subspaces means we can eliminate the process of automatic subspace segmentation as proposed by [7, 8]. It also allows us to dismiss the need for managing outliers or missing data in our subspaces [8].

3. System Overview

In order to create a given subspace, PCA is performed on the set of processed images for each hand shape. Processing involves segmenting hand objects from each image and scaling them to 32*32. Scaling is simply achieved by resizing the pixels contained within the objects bounding box. PCA provides M orthogonal eigenvectors {u1, … ,uM} of the covariance matrix, that correspond to the first M largest eigenvalues, in order to maintain a minimum energy of the dataset. In our experiments we have found that retaining 97% of energy is sufficient to differentiate hand shape subspaces.

The perpendicular distance (D_p) of a point **p** to a given eigenspace **E** can then be found using equation (**1**).

$$D_p{}^2 = D_e{}^2 - \sum_{i=1}^{M} [(\underline{p} - \underline{o}) \bullet \underline{u}_i]^2 \qquad (1)$$

Where D_e = Euclidean distance between **p** and the origin **o** of **E**.

We now have the backbone for a simple finger spelling recognition system. ASL contains 24 static hand gestures. A sample system can be developed as shown in Figure 1. It is constructed as follows:

Training - Generating a transformation subspace for each hand shape.

Testing – Project the test image into each of the subspaces to find the subspace with the nearest perpendicular distance. This subspace will be representative of one particular hand shape.

3.1. Reducing search time

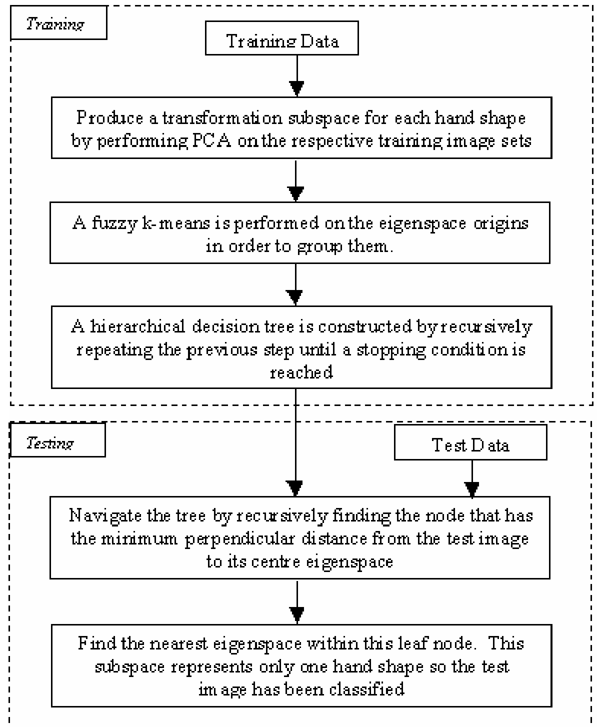


Figure 1. Reducing Search Time Overview

Initial experiments have shown that it is possible to reduce the search time from that of an exhaustive search. This can be achieved by organising the PCA reduced subspaces into a hierarchical decision tree. The decision tree can then be constructed using a fuzzy k-means algorithm that divides the dataset into two groups. This process is recursively executed until the stopping condition of the data in a node reaching a variance threshold is reached.

We have found that the origins of the eigenspaces are appropriate constituents to estimate comparability of eigenspaces. Consequently k-means is performed on the origins in order to group eigenspaces and create the search tree.

In a binary tree, at each stage the test image will have the option of following either path. The test image chooses the path by which it has a smaller perpendicular distance to the center eigenspace of that node. The center eigenspace of a node is estimated by, finding both the average origin and average eigenvectors of the eigenspaces contained in that node. We have found this approximation of the center image is sufficient for the test image to successfully navigate the tree.

Once a leaf node is obtained, the nearest subspace in that node to the test image can be found.

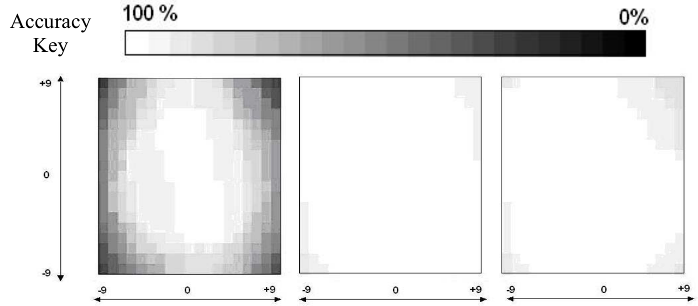
4. Experiments

We have performed a series of tests to assess the transformation subspace technique. We compare it with the Tangent Distance with relation to speed and accuracy. Code used for Tangent Distance is as found at [11].

All test and training images are obtained using computer animation. The transformation subspace has been trained only on the origin hand shape and origin hand shapes translated in all directions using combinations of 2, 4 and 6 pixels. Each subspaces is therefore created from 49 training images (7*7 manifold of translations in all directions). Translation occurs in the 380*380 image, produced by POSER, before scaling to 32*32. Because of the small number of transformation images used, each subspace can be sufficiently represented using 7 eigenvectors.

Figure 2 shows the performance images of the various algorithms. These images illustrate the results of testing with the 19 * 19 manifold of the 24 origin hand shapes images translated in all directions. Table 1 presents a subset of this test set. Here the test set is 96 (24 origin hand shapes, each translated up, down, left and right). It is evident that the tangent distance is only beneficial for small image translation transformations. Conversely the transformation subspace method presents 100% accuracy for the range of images it has been trained on. The Subspace Tree Search compares quite well to the regular subspace search. Also illustrated is the superior speed of the Tangent Distance. However, the subspace technique offers ample efficiency for the task of hand shape recognition in real time. The subspace tree search goes someway to addressing the speed issue but in doing so we compromise, slightly, on accuracy. Note all experiments are run on a standard PC using the matlab interpreter with non-optimised code.

Table 2 shows the results of training and testing with both translated and rotated object images. Test images contain the rotations as described in Table 2 along with translation combinations of 1,3 and 5 pixels. The test set for each rotation then contains 1176 images, 24 hand shapes at 49 differing translations.



(a) (b) (c)

Figure 2. Performance images for (a) Tangent Distance, (b) Subspace Distance and (c) Subspace Tree Distance.

Table 1. Comparison of the performance of distance metrics for translated images.

Distance algorithm	Recognition Rate (%) for test images translated by the following number of pixels									Average Speed Per Image (seconds)
	1	2	3	4	5	6	7	8	9	
Tangent	100	100	97	85	74	60	40	28	21	0.0042
Subspace	100	100	100	100	100	100	100	97	94	0.0084
Subspace Tree Search	100	100	100	100	100	100	99	96	92	0.0068

Table 2. Performance of Tangent Distance and Subspace Distance for translated and rotated object images.

Distance algorithm	Recognition Rate (%) for test images rotated by the following number of degrees and using different translations.									Average Speed Per Image (seconds)
	-12°	-9°	-6°	-3°	0	3°	6°	9°	12°	
Tangent	54	70	84	81	86	83	77	62	46	0.0042
Subspace	96	99	100	100	100	100	100	99	93	0.0098
Subspace Tree Search	95	99	98	97	96	95	95	93	90	0.0072

Table 3. Performance of Tangent Distance and Subspace Distance for translated and rotated object images that contain random shape distortions.

Distance algorithm	Recognition Rate (%) for test images rotated by the following number of degrees and using different translations along with ad-hoc shape distortion.									Average Speed Per Image (seconds)
	-12°	-9°	-6°	-3°	0	3°	6°	9°	12°	
Tangent	47	63	70	77	78	72	66	53	40	0.0044
Subspace	90	93	99	99	98	96	89	86	85	0.0104
Subspace Tree Search	87	87	94	95	96	94	88	85	85	0.0078

Each hand shape subspace has been created on a set of training images that contain translated origin images, translated origin images rotated 6° left and translated origin images rotated 6° right. As before only translations of 2,4 and 6 pixels are used.

Once again the subspace distance demonstrates a superior performance. The subspace Tree Search offers similar results, while the Tangent Distance technique deteriorates when non-trivial rotations are tested. A reason for the inferior performance is that the object image rotations are acquired in 3D space whereas Tangent Distance aims to approximate 2D image rotation transformations.

In table 3 the result of introducing random shape variations to the test images are displayed. Random variations are achieved by slightly deviating each joint in the hand in an unsystematic manner. The test set for each rotation is 3528, 3 random shape variations of 49 translations for each of the 24 hand shapes. The transformation Subspace was trained as in Table 2. Clearly the subspace technique and the subspace Tree search provides a better invariance to random hand shape variations.

These experiments show that while Tangent Distance is an effective technique for small 2D image transformations, its usefulness does not compare well with transformation subspace distance for robust hand shape classification.

5. Conclusions

In this paper, a robust hand shape recognition system has been proposed based on a subspace classifier. The subspaces are constructed, with a-priori knowledge, from images acquired using POSER modeling software. One important aspect of this approach is that once the allowable hand shapes and their bounds have been defined, the set of images of allowable transformation can then be automatically extracted without the expensive need of cameras and actors. This novel method also means our transformation subspace can be complete and free from outliers allowing for accurate robust recognition. Another important aspect is that the subspaces of 2D images are created from 3D transformations; this further enhances the accuracy of the recognition.

We also introduced a tree-search technique to improve the search time from an exhaustive search. It is hoped that future research will allow us correctly classify a large number of hand shapes from any permissible viewpoint.

6. Acknowledgments

This work has been funded by the Irish Research Council for Science Engineering and Technology (IRCSET)

7. References

[1] H. Wu, "Gesture Recognition Using Principle Component Analysis, Multi-Scale Theory, and Hidden Markov Models", *PHD Thesis*, Dublin City University 2002.

[2] W. Zhao, "Improving the robustness of subspace FR system" *2nd Int. Conf. Audio and Video Based Person Authentication, Washington DC*, pp 78-83, 1999.

[3] Atid Shamaie and Alistair Sutherland, "Accurate Recognition of Large Number of Hand Gestures," *2nd Iranian Conference on Machine Vision and Image Processing*, K.N. Toosi University of Technology, Tehran, Iran, 2003.

[4] Feng-Sheng Chen, Chih-Ming Fu, Chung-Lin Huang, "Hand Gesture Recognition Using a Real-time Tracking Method and Hidden Markov Models", *Image and Vision Computing,* Vol 21, pages 745-758, 2003.

[5] T. Kadir, R. Bowden, E. J. Ong, A. Zisserman, "Minimal Training, Large Lexicon, Unconstrained Sign Language Recognition", *Proceedings of the 15th British Machine Vision Conference, Kingston,* 2004.

[6] J. M. S. Dias, P.Nande, N. Barata, A. Correia, "O.G.R.E – Open Gestures Recognition Engine" *Procedings of the 17th Brazilian Symposium on computer Graphics and Image Processing 2004 (SIBGRAPI'04),* 2004.

[7] R. Vidal, Y, Ma, S. Sastry, "Generalized Principle Component Analysis (GPCA)" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 27, No 12, 2005.

[8] H. Bishoff, A. Leonardis, J. Maver, "Multiple Eigenspaces", *Pattern Recognition*, Vol 35, pp 2613-2627, 2002.

[9] P. Y. Simard, Y. A. Le Cun, J. S. Denker, B Victorri, "Transformation Invariance in Pattern Recognition – Tangent Distance and Tangent Propagation", Proc. Neural Networks: Tricks of the trade, Lecture Notes in Computer Science, vol. 1524,pp 239-274, Springer Verlag, 1998.

[10] Curious Labs, [Online], Available: http://www.e-frontier.com/, (Accessed 9 December 2005).

[11] D. Keysers Tangent Distance Implementation, [Online], Available: http://www-i6.informatik.rwth-aachen.de/~keysers/td/, (Accessed 9 December 2005).