Indian Institute of Technology Kanpur

CS 676 : Computer Vision and Image Processing

Articulated Hand Posture Recognition System using IDSC Features

Sourav Khandelwal

April 19, 2011

Abstract

In this project, I studied the application of inner distance shape context descriptor(IDSC) in recognizing articulated objects particularly the hand articulation. The IDSC is used as a hand shape descriptor to capture the hand state(open/closed) and to recognize between them. Further, the technique is also used to recognize sign languages. My analysis is based on the SVM classifier and contour hand images obtained from the Image and Video computing group of Boston University

Acknowledgements

Introduction

There are variety of hand postures possible with the articulated hand. The correct recognition of the different postures has been challenging task because of the large degrees of freedom(27 DOFs) of hand and the occlusion of fingers. The normal shape context proposed by Belongie et al. [1] in 2002, which is used to measure the similarity between shapes does not work well with complex articulated shapes. Inner distance proposed by Ling et al.[2], has been proved to captured the part structures of articulated shapes very well.

Shape Context

The shape context is a descriptor used to measure similarity and point correspondences between shapes. It describes a point on the object contour wrt to every other points in the contour. If given n points, $p_1, p_2, .., p_n$, on the object's contour, the shape context descriptor of a point p_i is a coarse histogram h_i of the relative coordinates of remaining $n - 1$ points.
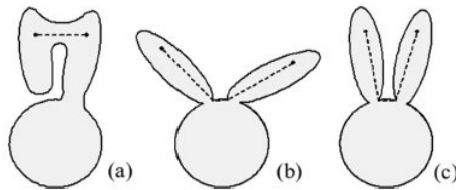
$$h_i(k) = \#\{q \neq p_i : (q - p_i) \in bin(k)\}$$

Here the bins uniformly divide the log-polar space. The computation of histogram is based on both distance and angle for each point on the contour with respect to all other points. This descriptor is a robust, compact and highly discriminative description of objects as it captures the distribution of each point relative to all other points in the object contour. However, it is not invariant to shape articulations as it uses L2 distance.

Inner Distance

The concept of inner distance was first proposed by Haibin Ling. Inner distance between two points is defined as the length of the shortest path within the shape boundary, to built shape descriptors **(IDSC)**. Inner distance has been proved to be invariant to shape articulations. Intuitively it can be seen from the following figure. Shapes (a) and (c) have similar spatial distribution and have quite different in their part structures. On the other hand, shapes (b) and (c) are the same objects with different articulations. The inner distance between the two points in (a) and (b) is different while for (b) and (c), it is similar. This shows that inner distance captures shapes and is invariant to shape articulations.

Figure 1: inner distance using points on shape boundary

1 Methodology

The overview of the hand posture recognition methodology is described here. The palm region is first extracted from hand images using skin based segmentation. Then, the longest contour is extracted from the segmented images using boundary extraction. 200 points are sampled uniformly from the contour and represented in the form of a graph array storing the euclidean distance between points.IDSC descriptor is calculated using these 200 points. This returns a histogram description of each point along the objects' contour to describe other points in the contour with respect to distance and angle.

The hand posture images contains a good amount in-plane rotations and out-of-plane rotations. Therefore, the images are passed through a pre-processing step where the train and test images are grouped based on their primary orientation direction. The training images are grouped uniformly into 10 intervals in 0-180 degree range. After the preprocessing is done, the IDSC descriptors are used as feature representation to train the SVM for the corresponding orientation. For testing, the orientation of the test image is found and is projected onto the appropriate SVM.
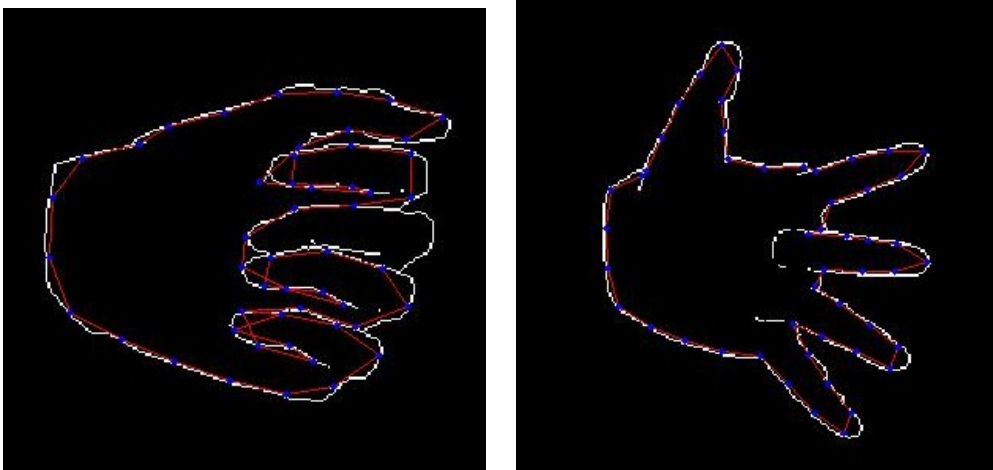
IDSC Computation

The IDSC is intuitively extended from the shape context. As in shape context, in IDSC, the computation of histogram is based on both distance and angle for each point on the contour with respect to all other points. In shape context, the distance between points is the normal L2 distance where as IDSC uses inner-distance calculated as the length of the shortest path within the shape boundary. The angle is used to measure the relative orientation between points. In case of IDSC, inner-angle is used. The computation of inner-distance and inner-angle is described below.

Inner Distance Computation

In order to find the inner distance, we need to consider the length of the shortest path within the shape boundary. Therefore, first we extract the longest boundary of the object and sampled n contour points, $p_1, p_2, .., p_n$ on the boundary. Now, the inner-distance is computed using shortest path algorithm. A graph is built using sampled points. For each pair of sample points p_1 and p_2 , if the line segment connecting p_1 and p_2 falls entirely within the object, an edge between p_1 and p_2 is added to the graph with its weight equal to the Euclidean distance $||p_1 - p_2||$. An example is shown in figure 2. Once the graph is built, the inner distance matrix is computed using Bellman-Ford's all pair shortest path algorithm.

Figure 2: Graphical representation of boundary points for inner-distance

Inner Angle Computation

Inner-angle between two points p and q is calculated as the angle between contour tangent at p and the direction of path (p,q). This is called inner-angle. The inner-angle is used for orientation bins. This is used to make the descriptor rotation invariant. The figure shows the calculation of inner-angle.
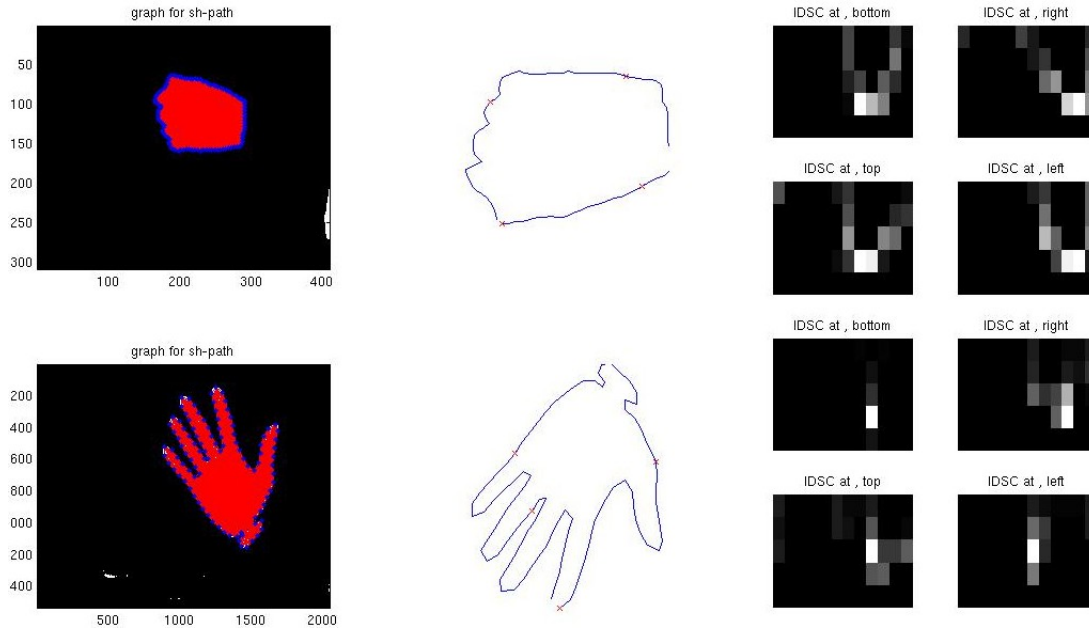
Figure 3: Inner-angle between p and q

IDSC Representation

The following image shows the idsc histogram representation for two images, one for closed hand and one for open hand. The left column is the graph formation using the boundary sampled points. The middle column represents the contour boundary. The right column represents the histogram at four points in the boundary, bottom-left, bottom-right, top-right, top-left.

Figure 4: Representation of IDSC

Hand Orientation Estimation

The orientation of the hand is found by calculating the primary scatter direction of the image. The scattering is estimated through principal component analysis (PCA). PCA solves the generalized eigen-value problem and computes the eigenvectors from the covariance matrix of the input images. The eigenvectors (corresponding to large eigenvalues) represents the directions of maximum scatter of the data. So, given a segmented hand image, we can estimate its scatter direction based on the co-ordinates of the eigenvector that has the maximum eigenvalue. An example of the scatter direction calculation is shoen in figure 5.
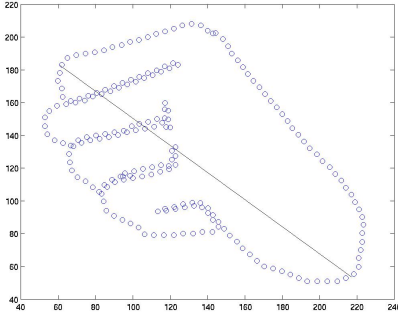
4

Figure 5: Scatter direction through PCA

Classification using Support Vector Machines

SVM is a standard statistical technique well known for classifying two-class problem. The SVM's are trained using labeled IDSC descriptors which are belonging to open and closed hands. SVM tries to find out a linear hyper plane which separates the two classes.

2   Experiments

(A) *Hand State Recognition*

The SVM is trained using IDSC descriptors of open/closed hand images obtained from Image and Video computing group, Boston University. About 100 images from each state are used. 200 contour points are sampled from the boundary to be used for the IDSC histogram representation. The histogram is represented in log polar space using 5 distance bins and 12 angular bins. The hand images contained a good amount of in-plane and out-of-plane rotations. About 100 images from each state are used for testing. The IDSC descriptors are obtained for the test images and projected onto the trained SVM. Some sample images used for this experiment is shown in the figure below. Using normal SC descriptors which uses euclidean distance between pair of points, the recognition was 64%. 74% recognition is obtained using IDSC without the pre-processing step. Using the pre-processing step, the accuracy has been increased to 81%. The confusion matrix for the recognition with pre-processing step is shown in table 1.
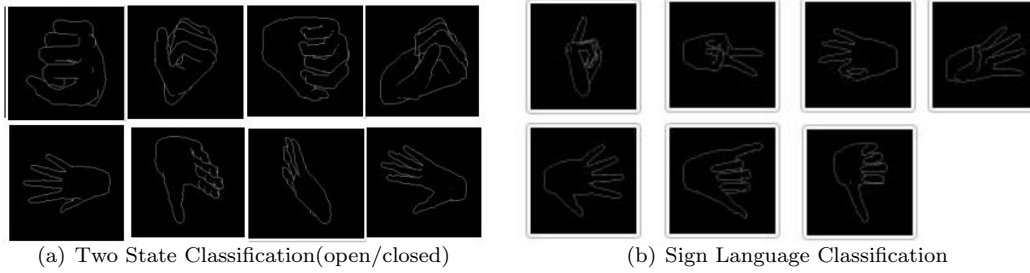
I have tried this experiment with my own hand dataset of open and closed images. Closed and open hand images are taken for 5 different persons using Microsoft's VX 3000 webcam. There were about 16 hand images of each person with 8 images in each state(open/closed). Leave-one-out technique is used to calculate the accuracy of the recognition. The accuracy obtained in this case is only 59%.

Table 1. *Hand State Recognition: Confusion Matrix*

	closed	open
closed	68	32
open	6	94

(B) *Sign Language Pattern Matching*

In this experiment, I used about 7 sign languages in which the hand represents one, two, three, four, five, call, and thumbs. For recognition, multi-class SVM is used. About 50 images for each sign language is used to train the multi-class SVM. This multi-class SVM uses the technique of one-against-all configuration to classify the images. The kernel for SVM used in this case is Gaussian. Other parameters, contours points, histogram configuration, etc, remains the same. The test dataset were remaining images of each sign languages obtained from the dataset of Boston University. Test images are about 59 for *one,*

(a) Two State Classification(open/closed)   (b) Sign Language Classification

57 for *two*, 47 for *three*, 64 for *four*, 76 for *five*, 48 for *call*, 74 for *thumbs*. Some sample images are shown in the figure below.. This experiment showed about 67% recognition. The confusion matrix is shown in the table 2.

Table 2. *Sign Language Pattern Matching-Confusion Matrix*

	one	two	three	four	five	call	thumbs
one	16	11	1	1	5	0	25
two	0	38	0	0	2	1	16
three	0	1	27	1	4	5	9
four	1	2	2	37	9	1	12
five	0	4	2	4	62	0	4
call	1	0	0	0	2	41	4
thumbs	0	4	0	1	6	1	62

3 Analysis

A brief description of the salient features that can be inferred from the above experiments is as follows:-

- The results obtained using inner-distance shape context descriptor has been found to be more accurate than normal shape context descriptor as seen in the case of hand state recognition.

- The confusion matrix of hand state recognition shows that there are many images of closed hand misclassified into open state. These images corresponds to the partially closed hand postures. Also, some of the misclassified images corresponds to those hand posture images in which the fingers were occluded.

- For the sign language pattern matching experiment, the confusion matrix shows that many hand images of state *one* and *two* are misclassified into the *thumbs* posture. This is because of their same articulated shape structure.

- The experiment of hand state recognition conducted on open and close d hand images obtained from webcam did not show good results. The possible reasons could be that the experimental setup for taking the images may not be accurate and the contours obtained from these hand images lagged the details of shape boundary description.

4 Conclusion

My aim in the undertaking of this study, was to appreciate the approach of inner-distance to capture articulated shapes and recognize them correctly. The project that I have done, has allowed me to apply various image processing and machine learning techniques in the domain of articulated objects, and to answer the important question pertaining to the recognition of hand postures.

I have dealt with two scenarios of hand posture recognition in this project. The experiment showed a good accuracy in recognizing the articulated hand postures using *idsc* in both of them. The idea can be extended to real time hand posture recognition from videos. The recognition could have been more effective if the descriptor could also capture the 3D view of the hand image as many cases of misclassification were from the hand posture images of occluded fingers.

Keeping in mind the fundamental objective of this experiment, which has been to investigate the ability to recognize articulated hand postures - this has been an encouraging step forward. It is still however too early to generalize this to all the articulated objects, because I have been selective in the data which is being analyzed, as well as the number of classes we are trying to categorize the data into. Lastly, the confluence of machine learning and modern image processing techniques, has been exciting to observe. Undoubtedly, the approach will continue to evolve.

5 References

[1] S. Belongie, J. Malik and J. Puzicha. "Shape Matching and Object Recognition Using Shape Context," IEEE Trans. Pattern Anal. Mach. Intell., $24(24) : 509 - 522, 2002$.

[2] H. Ling and D.W. Jacobs. "Shape Classification Using the Inner-Distance". IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI), pages 286–299, 2007.

[3] Gopalan, Raghuraman and Dariush, Behzad, "Toward a vision based hand gesture interface for robotic grasping", IEEE/RSJ international conference on Intelligent robots and systems, pages $1452 - -1459, 2009$.

[4] Hand Image Dataset, Image and Video computing group, Boston University.

[6] *MATLAB*, Language for Scientific Computing

[5] IDSC descriptor, Matlab code, Haibin Ling homepage.

[6] Multi-class SVM toolbox, Matlab.