Actuation in Perception: Character Classification in Engineering Drawings

Thomas C. Henderson[1], Narong Boonsirisumpun[2], and Anshul Joshi[3]

***Abstract*— We propose a novel approach to 2D character recognition by incorporating actuation data into the shape representation. Sensorimotor data is analyzed in terms of actuation sequences which generate the data. We illustrate the use of Wreath Products (WPs) to represent robot sensorimotor experience in a way that ties together perception and actuation. WPs naturally represent not only the Euclidean symmetries possessed by an object, but also the sequence of actuations used to generate those. Two distinct approaches using actuation signals to represent shape are compared: (1) the Kullback-Leibler measure is applied to histograms of translation symmetries in the shape, and (2) a distance metric is defined on pure actuation signals. Experimental results show that these methods achieve excellent classification rates (99 %) on text extracted from scanned images of engineering drawings for the top five hypotheses.**

I. INTRODUCTION

Most shape representation methods work directly in terms of features of the geometry of the particular shape. That is, a collection of 2D or 3D points is segmented from a 2D or 3D image or point cloud dataset, and then shape invariants are determined which uniquely identify the shape. Such features may or may not allow the recovery of the original set of points (see [2]). We have been exploring the use of the wreath product (WP) group as defined by Leyton [7] as a means of shape and concept representation. This approach is different from standard methods in that it defines action processes which generate the point set constituting the shape. These actions are defined as symmetry groups (e.g., translation, rotation, reflection, etc.) and a hierarchical shape representation results; this is achieved by having groups act on other groups (e.g., translate a point, rotate a set of points, etc.) (for more details see [3], [4], [5], [6]). For example, a square results from taking a point and translating it along a line for some distance, then rotating this line segment by 0, 90, 180 and 270 degrees to form the square. This inclusion of action processes in the shape description allows for straightforward knowledge transfer to a variety of actuation systems.

One major issue with this approach is that the symmetry groups (operators) are defined in some specific Cartesian coordinate frame; i.e., a translation is along a line defined in that frame. One of our goals in this work is to find a representation in terms of the natural motor actuation signals of the observing robot agent. This leads to our hypothesis:

> Shape representation based on the generative actuation process of the observing agent can be effective and efficient.

This is the basic idea explored in this paper; we propose a shape recognition method based on encoding the shape in terms of the actuation signal needed to generate (or observe) the shape. The method is effective in that it robustly characterizes shape, and it is effective in that it has low computational complexity.

II. RELATED WORK

In a more general setting, autonomous robots are embedded in some static or dynamic environment, and are expected to represent and carry out tasks in this environment in the presence of sensing and actuation uncertainty. This requires the agent to have a robust representation of its environment, as well as of the plans and actions that it can execute in that environment [8]. Our work here deals with creating such a representation that is robust by (1) being abstract in nature and grounded in (symmetry) theory so that it is more general, and (2) being mathematically well-structured so that it is practical. Leyton [7] hypothesized that wreath products form the basis for cognition by giving a generative representation of shape and structure, which not only gives the Euclidean symmetries present in a shape, but also encodes the actuation sequences (sequences of symmetry operations) that are used to generate a shape; however, he gave no details for a practical implementation. These representations are also based on a group theoretical concept, the Wreath Product, that is a special form of a semi-direct product [1], and lends itself to a practical implementation. We propose to validate this theory by creating a framework that constructs and exploits WP structures as the basis for robot cognition. Such an implementation needs processes for WP discovery from sensor and actuator signals, storage and retrieval of WPs, converting WPs to plans, and structured knowledge transfer within a single or multi-robot system. We have proposed a framework for implementing the WPs and demonstrate their effectiveness at representing shape and structure, the symmetries and actuations they encode, as well as spatial relations between objects. This framework is exploited to classify 2D characters.

Here we extend our previous work [6] to include direct incorporation and exploitation of actuation data in the analysis of shape. Note that Noë [9] provides a philosophical and

[1]Thomas C. Henderson is a full professor with the School of Computing, University of Utah, UT 84112, United States tch@cs.utah.edu

[2]Naron Boonsirisumpun is a PhD student with the School of Computing, University of Utah, UT 84112, United States narong.boonsirisumpun@utah.edu

[3]Anshul Joshi is a PhD candidate with the School of Computing, University of Utah, UT 84112, United States joshi@cs.utah.edu

psychological argument for the primary role of actuation in perception. He states (p. 102):

> The sensorimotor dependencies that govern the *seeing* of a cube certainly differ from those that govern the *touching* of one, that is, the ways cube appearances change as a function of movement is decidedly different for these two modalities. At an appropriate level of abstraction, however, these sensorimotor dependencies are isomorphic to each other, and it is *this* fact – rather than any fact about the quality of sensations, or their correlation – that explains how sight and touch can share a common spatial content. When you learn to represent spatial properties in touch, you come to learn the transmodal sensorimotor profiles of those spatial properties. Perceptual experience acquires spatial content thanks to the establishment of links between movement and sensory stimulation. At an appropriate level of abstraction, these are the same across the modalities.
>
> We can illustrate this by means of a simple example. If something looks square, then one would need to move one's head in characteristic ways to look at the corners. One would have to move one's hands *the same way* at the appropriate level of abstraction to feel each corner.

Note that Noë's discussion may involve more of the human actuation system (e.g., neck, torso, etc.) than we exploit in the character recognition problem.

III. WREATH PRODUCT CONSTRAINT SETS

In order to understand a wreath product we briefly explain the concept of a *semidirect product* in group theory, that underlies the concept of a wreath product. Consider a homomorphism ϕ given by $\phi_h(n) = hnh^{-1}$ for all $n \in N$ and $h \in H$, where H and N are groups, and H is a group that acts on N by conjugation. For each $h \in H$, conjugation by h is an element of $Aut(N)$ (automorphism group of N).

Given two groups N and H, and a group homomorphism ϕ from H into $Aut(N)$, $N \rtimes_\phi H$ denotes the *semidirect* product of N and H with respect to ϕ and satisfies the following:

1) $N \rtimes H$ contains elements from $N \times H$
2) Group operation of $N \rtimes H$ is defined as: $(n_1, h_1)(n_2, h_2) = (n_1\phi_{h_1}(n_2), h_1h_2)$, where $n_1, n_2 \in N$ and $h_1, h_2 \in H$.

Now consider a group L where L consists of the direct product of $k =\mid H \mid$ copies of N, i.e., $L = N_1 \times N_2 \times ...N_k$. **The *wreath product*, $\mathbf{G = N \wr H}$, is formed by the *semidirect product* of L and H. Thus $\mathbf{G = N \wr H \equiv G = L \rtimes H}$.**

We propose *Wreath Product Constraint Sets* (WPCS) as a mechanism to represent shape (here: lower and upper case English letters and the digits 0 to 9). A WPCS:

1) Uses $\Re$ and $O(2)$ wreath product groups as the basic shape constituents.

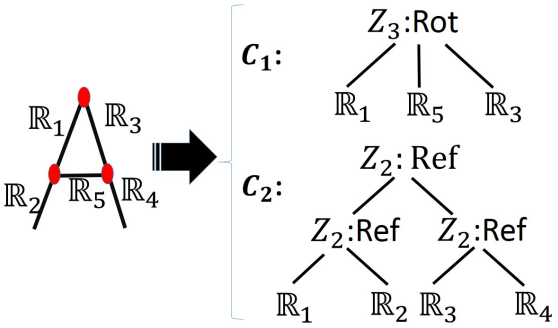

Fig. 1. Wreath Product Constraint Set for Letter 'A'

2) Enumerates further wreath products that hold between constituents (even singletons).
3) Adds specific (geometric) constraints between shape constituents (e.g., set operations).

Note that we use $\Re$ to represent the 1D translation symmetry group, $O(2)$ for the 2D rotation symmetry group (both of these are continuous), and Z_n to represent the cyclic group of order n (e.g., discrete set of rotations). As an example, consider the representation of the upper case letter 'A' shown in Figure 1. The left side of the figure shows the basic constituents of the letter 'A' – in this case five $\Re$ groups; the right hand side of the figure shows the two constraints in the WPCS: (1) C_1 describes the triangle in the letter, and (2) C_2 describes the reflection symmetry between the left and right side line segments. Note that Z_2 is the cyclic group of order 2 and models several geometric symmetries. We denote reflection by adding the annotation : Ref, and rotation by : Rot. Each of these groups has its own specific coordinate axes (e.g., the $z-$axis for rotation, and a specific line in the plane for the reflections. For the top-level Z_2 : Ref group in C_2 this axis is the $y-$axis, while for the lower level reflections, it will be the line bisecting the respective side at the points indicated in the drawing (on the left of the figure). R_i in the figure is the i^{th} straight line segment. Note that there will be additional information added to the representation to describe the actuation processes which give rise to these constituents (see below). Compare this to the WPCS representation of the capital letter 'H' shown in Figure 2. It can be seen that there is only one constraint in the set (the triangle is not found), and the highest level reflection symmetry describes the horizontal reflection of the entire 'H' figure. Thus, the WPCS representation exposes both the similarities (e.g., the common subgraph) between the two letters, as well as the differences. Also, note that there are multiple WPCS representations for a set. For example, the letter 'A' can also be represented as the two sides and the cross bar in the middle (i.e., 3 $\Re$ groups, instead of 5).

The basic WPs for letter representation are $\Re$ (straight line segments) and $O(2)$ (circles). Therefore, we have developed special analysis algorithms to produce $\Re$ and $O(2)$ hypotheses.

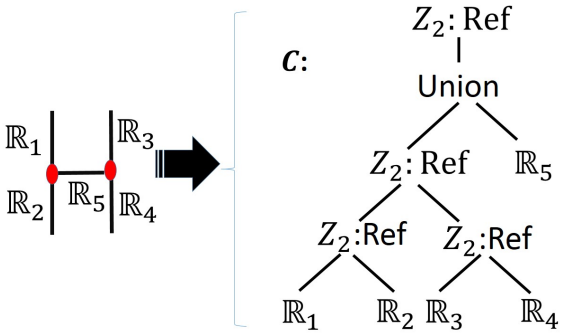

Fig. 2. Wreath Product Constraint Set for Letter 'H'

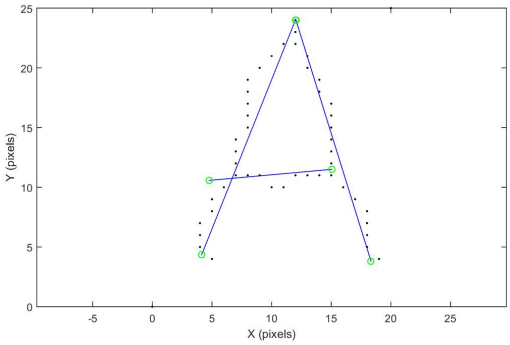

Fig. 3. $\Re$ Hypotheses for Letter 'A'

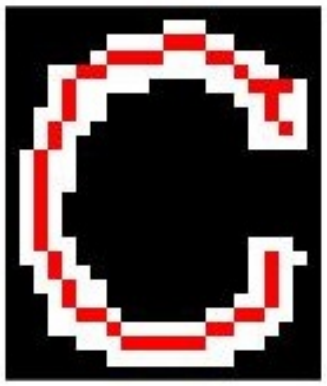

Fig. 4. Skeleton overlaid on Original Image for Letter 'C'

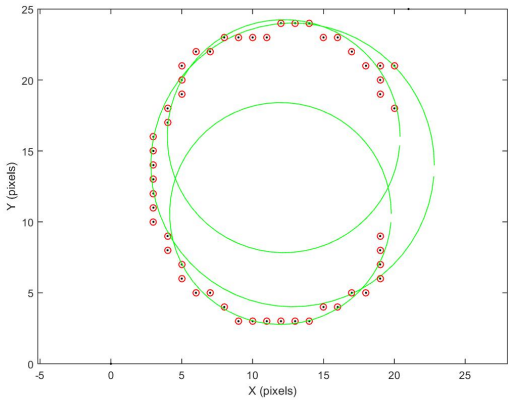

Fig. 5. $O(2)$ Hypotheses for Letter 'C'

A. $\Re$ Hypotheses

In order to discover $\Re$ hypotheses, we use the connected component image and its skeleton image (i.e., medial axis transform). Figure 8 shows a skeleton overlaid on the original image. The basic logic of the $\Re$ hypothesis approach is:

```
for every pixel, p, in the skeleton
  V := skeleton pixels visible from p
  R hyp := pixel sets with p as endpt
```

Figure 3 shows the $\Re$ hypotheses found in a sample capital letter 'A' skeleton image. V is formed by checking visibility in terms of the a straight line of pixels in the original image connecting p and the visible skeleton pixel. p is an endpoint if it is one of the two most distant points of the projection of the points in V onto the best fit line to the points in V.

B. $O(2)$ Hypotheses

Circular sections (or parts thereof) are more difficult to find. Figure 4 shows the letter 'C' and its skeleton. The basic logic of the $O(2)$ hypothesis approach is:

```
for every pixel,p, in the skeleton
  flow := distance of pixel from p
  T := pixels within distance 15 of p
  O(2)_hyp := best fit circles in T
```

Figure 5 shows the $O(2)$ hypotheses found for the sample letter 'C' image. As another example, Figure 6 shows the lower case letter 'a' image, and Figure 7 shows the $O(2)$ hypotheses found for it. Distance here is in terms of chain code distance (8-neighbor steps).

C. Character Templates and Segment Classification

Given $\Re$ and $O(2)$ basic constituents, it is possible to define WPCS templates for all the character shapes; e.g., those for the letters 'A' and 'H' shown above. It is also possible to learn the templates by taking a set of training samples, extracting the constituents and finding the constraints between the constituents. This would involve either setting a hard threshold to produce predicates for the constraints (e.g., for when a reflection holds between two point sets), or adding a probabilistic framework. In the course of this study, we discovered that the actuation signals which encode the basic constituents provide an effective and efficient shape representation for the WPCS; this is described in the next section.

IV. Actuation Signals as Representation

Since we do not have an embodied agent in this application, we resort here to *virtual actuators*, and in particular, a *virtual camera* for image acquisition and a *virtual hand*

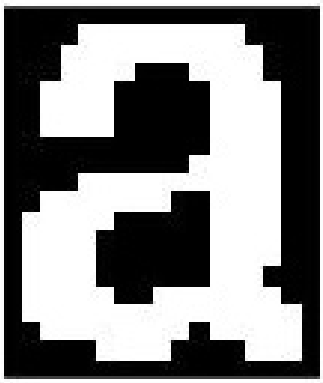

Fig. 6. Image of Lower Case Letter 'a'.

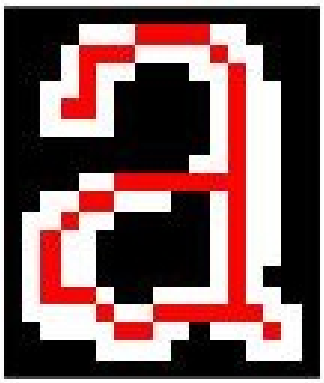

Fig. 8. Skeleton of Lower Case Letter 'a' overlaid on Original Image.

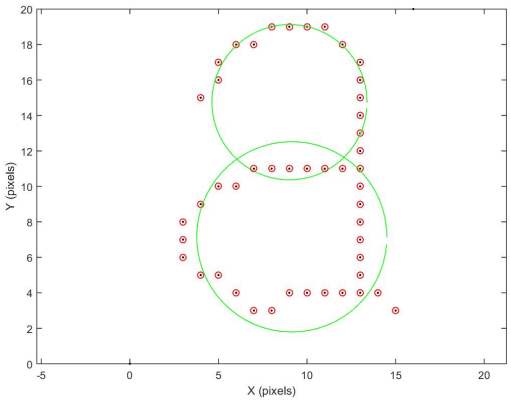

Fig. 7. $O(2)$ Hypotheses for Letter 'a'

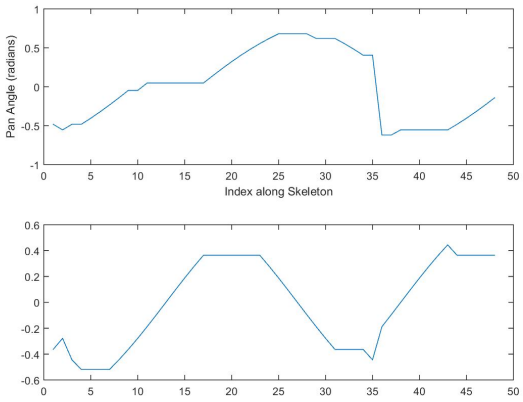

Fig. 9. The *pan* and *tilt* Signals for Lower Case 'a'

for shape generation. We show that a character can be represented as a wreath product constraint set which provides an abstract representation of the shape and allows for

1) recognition and classification of text characters, and
2) structured knowledge transfer from the camera actuation system to the robot hand actuation system in order to achieve shape synthesis.

A. Virtual Camera

Since we are working with images that have already been scanned, we have developed *virtual actuators* and corresponding actuation command streams for the given data acquired from the image. This works as follows:

- Each individual connected component is set in a circumscribing rectangle (Figure 6 shows the subwindow for the lower case letter 'a').
- A virtual camera is positioned in the middle of the image and above the image by one-half the length of the longer rectangular side.
- The skeleton of the connected component is found next (see Figure 8).
- The camera is aimed at each of its constituent pixels in turn, and the pan and tilt angles are recorded. E.g., Figure 9 shows the pan and tilt angles for the lower case letter 'a'.

B. Virtual Hand

The specific *virtual hand* considered here is an RR robot (i.e., two revolute joints). The base of the arm is located at the center of the character sub-window, and the lengths of the links are equal and set to one-fourth the length of the greater diagonal of the sub-window (see Figure 10). This allows the virtual hand to place its endpoint anywhere within the sub-window. Given that a shape will need to be generated in a standard Cartesian coordinate frame, it will be necessary, in general, to learn the transform from (pan,tilt) space to $(\theta 1, \theta 2)$ space so as to obtain the same (x, y) point. However, in this specific case, the transform is given as:

$$c_2 = D = \frac{(dtan(\theta_{pan}))^2 + (dtan(\theta_{tilt}))^2 - 2a^2}{2a^2}$$

$$s_2 = \sqrt{1 - D^2}$$

$$\theta_2 = atan2(s_2, c_2)$$

$$\theta_1 = atan2(dtan(\theta_{tilt}), dtan(\theta_{pan})) - atan2(s_2, 1 - c_2))$$

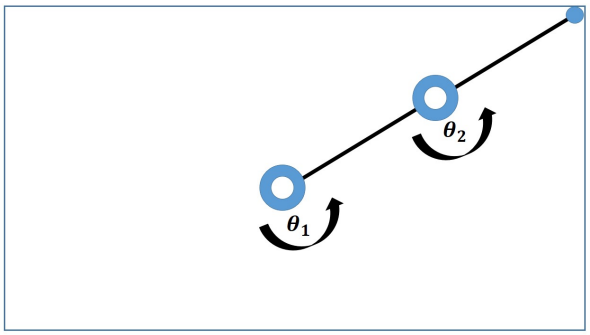

Fig. 10. A Simple 2-Revolute Joint (RR) Robot Hand.

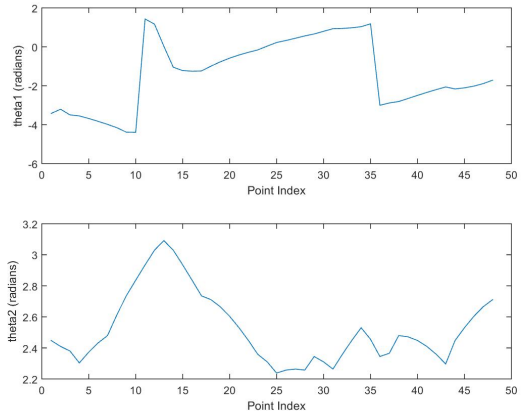

Fig. 11. Lower Case Letter 'a' Representation by (θ_1, θ_2).

E.g., the (θ_1, θ_2) signal representation for lower case 'a' is shown in Figure 11. In the experiments described below, we use the (pan,tilt) representation of the WPCS constituents.

V. Classification Based on Translation Symmetry

It is possible to represent shapes in terms of just the local translation symmetries. This means that for every pixel in the shape, the maximal translation direction is determined; this is done by finding the maximal set of pixels forming a line through the given pixel and that stay inside the shape. It is not necessary to recover the $\Re$ or $O(2)$ constituents for this approach. Statistics of the pixel-based translation symmetries provide enough information for shape classification. For experiments we use segments from the image shown in Figure 12. There are 1174 legitimate characters (although seven of these do not satisfy the classification assumptions – e.g., wrong Euler number, etc.); of these 1161 are correctly classified in the top five hypotheses (we allow multiple hypotheses and then reduce them when words are found).

Consider a character such as that shown in Figure 6. At each pixel, the direction and extent of the 1D translation is determined. Figure 13 shows the results for this character. Our first classification approach exploits the first order

2.6 Glass Reinforced Plastic – Exterior.

2.6.1 Clean with a lint free cloth dampened with thinner, type optional, Spec. TT-T-306.

2.6.2 Mechanically abrade surface using medium grit (180 to 220) media, or 180 to 220 grit sandpaper.

2.6.3 Clean surface with a lint free cloth dampened with thinner, type optional, Spec. TT-T-306.

2.6.4 Apply epoxy primer coating, lead and chromate free, Spec. MIL-P-53030, or epoxy primer coating, lead and chromate free, Spec. MIL-P-53022.

2.6.5 Apply polyurethane, type optional, Spec. MIL-C-46168, color green 383, or aliphatic polyurethane, single component, Spec. MIL-C-53939, color green 383.

2.6.6 Inspect per paragraph 5.0.

2.7 Titanium Alloy

2.7.1 Clean per method II, treat per type III, Spec. TT-C-490.

2.7.2 Apply primer coating, epoxy, lead and chromate free, Spec. MIL-P-53030, or primer coating, epoxy, lead and chromate free, Spec. MIL-P-53022, comp. optl.

2.7.3 Apply aliphatic polyurethane, type optional, Spec. MIL-C-53939, color green 383, or apply polyurethane, type optional, Spec. MIL-C-46168, color green 383, or apply modified high solids polyurethane, color green 383, with performance and color requirements of MIL-C-46168.

2.7.4 Inspect per paragraph 5.0.

3.0 Interior Paint System

3.1 Ferrous Metals (Rockwell Hardness less than C40) – Interior. (See Paragraph 4.0 for Coating Sequence.)

3.1.1 Clean per applicable methods I thru VI. Treat per type I or III, Spec. TT-C.490.

3.1.2 Optional: Apply primer coating, lead and chromate free, Spec. MIL-P-53030, or epoxy primer coating, lead and chromate free, Spec. IL-P-53022.

Fig. 12. Image used in Experiments.

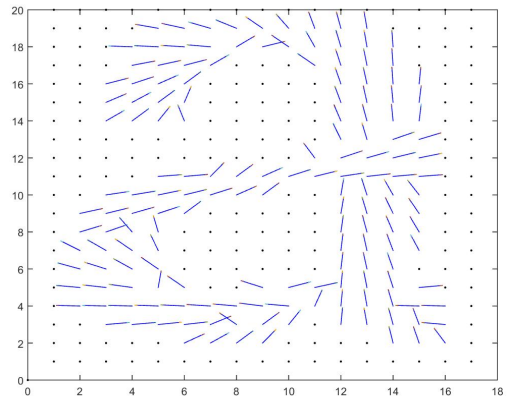

Fig. 13. Translation Symmetry Directions for Lower Case Letter 'a'.

statistic of orientation. The directions are aligned with the closest of 0, 45, 90, and 135 degrees.

In order to constrain the orientations somewhat according to their 2D distribution in space, we construct a histogram for each of four sub-regions of the image: (1) upper portion, (2) lower portion, (3) left portion, and (4) right portion (in this study the portion is 2/3's). The method also uses Euler number and horizontal and vertical symmetry measures. Figure 14 shows the four histograms catenated, and for comparison, the histogram from another letter 'a' from a test image. In practice the four histograms from a test character, and the norm of the resulting 4-tuple is used as the distance measure (in this case, the vector was $p = [0.0425, 0.0409, 0.0459, 0.0069]$). Each unknown character (connected component) in a test image is compared to each character template, and the top five matches are kept. Using this approach, we obtained a 95% classification rate;

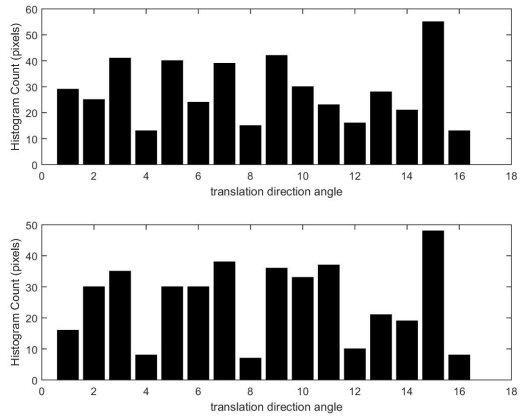

Fig. 14. Comparison of Combined Symmetry Translation Direction Histograms from Four Subwindows in Two Different 'a' Segments.

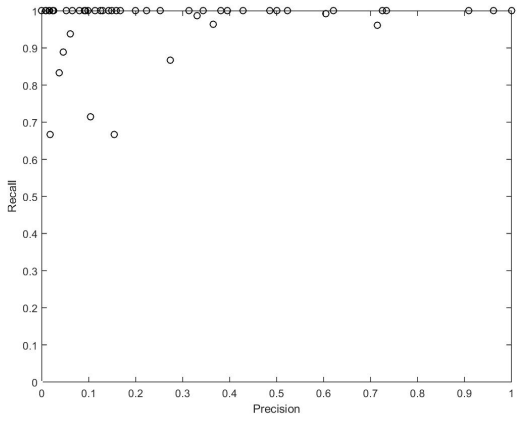

Fig. 15. Precision and Recall Plot for Symmetry Translation Classification.

when only the top 1, 2, 3, and 4 hypotheses are kept, the classification rate are 80%, 90%, 95% and 98%, respectively. Figure 15 shows the precision and recall plot for this method.

VI. Classification Based on Pan-Tilt Actuation signals

It is also possible to use explicit actuation data to classify unknown characters. Given a set of pixels along the skeleton of a character, the pan and tilt angle for the *virtual camera* are found as described earlier. For the template character 'a' the pan and tilt angles are shown in Figure 9. A distance measure between the shapes is then based on the difference of the two (pan,tilt) signals for the different characters. We use the following simple measure:

$$\mu((p_1,t_1),(p_2,t_2)) = \sum_{i\in\mathcal{I}_1} argmin_{j\in\mathcal{I}_2}(\mid (p_i,t_i)-(p_j,t_j) \mid)$$

where $\mathcal{I}_1$ and $\mathcal{I}_2$ are the index sets for points in shape 1 and 2, respectively. This is the sum of the distances to closest (pan,tilt) pair in the other set. Using this method, we obtained a 99% classification rate for the top five hypotheses;

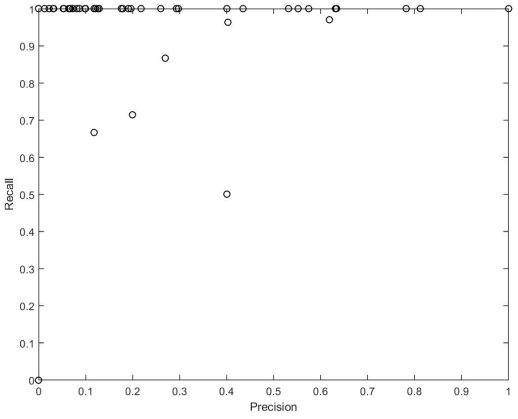

Fig. 16. Precision and Recall Plot for Pan-Tilt Actuation Data Classification.

when only the top 1, 2, 3, and 4 hypotheses are kept, the classification rate are 88%, 96%, 97% and 98%, respectively. Figure 16 shows the precision and recall plot for this method.

VII. Classification Based on RR Robot Actuation signals

We also explored the use of explicit 2-revolute joint robot actuation data to classify unknown characters. Given a set of pixels along the skeleton of a character, the two joint angles for the *virtual robot hand* are found as described earlier. For the template character 'a' the pan and tilt angles are shown in Figure 9. A distance measure between the shapes is then based on the difference of the two (θ_1,θ_2) signals for the different characters. We use the following simple measure:

$$\mu((p_1,t_1),(p_2,t_2)) =$$

$$\sum_{i\in\mathcal{I}_1} argmin_{j\in\mathcal{I}_2}(\mid (\theta 11_i,\theta 12_i)-(\theta 21_j,\theta 22_j) \mid)$$

where $\mathcal{I}_1$ and $\mathcal{I}_2$ are the index sets for points in shape 1 and 2, respectively. This is the sum of the distances to closest (θ_1,θ_2) pair in the other set. Using this method, we obtained a 98% classification rate for the top five hypotheses; when only the top 1, 2, 3, and 4 hypotheses are kept, the classification rate are 79%, 93%, 97% and 98%, respectively. Figure 17 shows the precision and recall plot for this method.

VIII. CONCLUSIONS AND FUTURE WORK

We have proposed using an actuation based representation of shape and demonstrated its effectiveness on a character classification problem useful in engineering drawing analysis – in this case for text images in the engineering drawing image set. In future work, we intend to perform more experiments on larger and more varied datasets to better characterize the strengths and weaknesses of the method. In addition, we intend to explore the use of the full WPCS representation to help reduce the error rate. For example, the

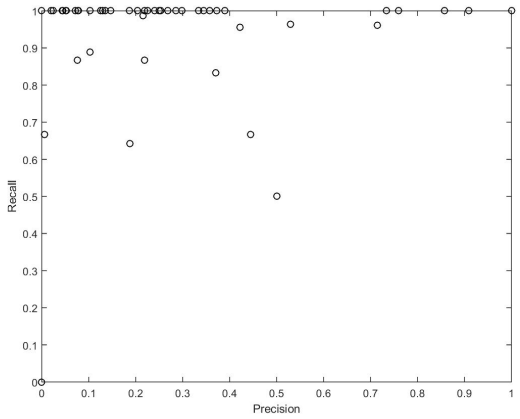

Fig. 17. Precision and Recall Plot for RR Robot Actuation Data Classification.

lower case letter 't' is sometimes confused with the lower case letter 'i' because the cross bar is not identified; this can be determined if a lower threshold is used for the $\Re$ hypotheses in the analysis of the skeleton. We also intend to extend this work to 3D shapes.

ACKNOWLEDGMENT

This work was supported by AFOSR grant FA9550-12-1-0291.

REFERENCES

[1] D.S. Dummit and R.M. Foote. *Abstract Algebra*. Wiley & Sons, Hoboken, NJ, 2004.

[2] D.A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Pearson, Boston, MA, 2012.

[3] T.C. Henderson, N. Boonsirisumpun, and A. Joshi. Symmetry Based Semantic Analysis of Engineering Drawings. In *Proceedings of the Conference on Multisensor Fusion and Integration for Intelligent Systems*, Beijing, China, September 2014. IEEE.

[4] T.C. Henderson, E. Cohen, E. Grant, A. Joshi, M.T. Draelos, and N. Deshpande. Symmetry as a Basis for Perceptual Fusion. In *Proceedings of the Conference on Multisensor Fusion and Integration for Intelligent Systems*, Hamburg, Germany, 2012. IEEE.

[5] T.C. Henderson, H. Peng, C. Sikorski, N. Deshpande, and E. Grant. Symmetry: A Basis for Sensorimotor Reconstruction. Technical Report UUCS-11-011, The University of Utah, May 2011.

[6] A. Joshi, T.C. Henderson, and W. Wang. Robot Cognition using Bayesian Symmetry Networks. In *Proceedings of the Internaitonal Conference on Agents and Artificial Intelligence*, Angers, France, 2014. IEEE.

[7] M. Leyton. *A Generative Theory of Shape*. Springer, Berlin, 2001.

[8] H. Liu, D. Gu, R.J. Howlett, and Y. Liu. *Robot Intelligence: An Advanced Knowledge Processing Approach*. Springer, Berlin, Germany, 2010.

[9] A. Noë. *Action in Perception*. MIT Press, Cambrideg, MA, 2004.