

A Developmental Framework for Visual Learning in Robotics

Amol Ambardekar¹, Alireza Tavakoli², Mircea Nicolescu¹, and Monica Nicolescu¹

¹Department of Computer Science and Engineering, University of Nevada, Reno, Reno, Nevada, USA

²Department of Computer Science, University of Houston –Victoria, Victoria, Texas, USA

Abstract - In this paper we are investigating a developmental learning strategy for robotic applications. Two approaches based on Isomap dimensionality reduction and on feature-based learning are investigated. In the training phase, the robot is presented with objects (e.g. squares, circles) and labels regarding their properties (e.g. shape, size, orientation). From these the system learns a representation of each of these properties (concepts) and is then able to recognize them in new images, previously unseen. Based on our experiments, we have concluded that in the Isomap space, given enough samples of objects and their properties, each dimension represents one or a combination of the object properties. The feature-based approach uses a pre-processing technique to extract pre-defined features from an object image. Our main investigation is concentrated in the Isomap space and related to finding an automated inference mechanism for learning object properties (concepts).

Keywords: Learning strategies, computer vision, robotics, clustering techniques, image feature extraction.

1 Introduction

Young children learn the conceptual meaning of objects by associating auditory and visual sensory information they receive about them. As the young children and toddlers explore the world and interact with adults they populate their repertoire of conceptual information related to objects' features, such as object colors, shapes, sizes, etc. When interacting with adults, the children may learn about objects by acquiring information from their audio-visual sensory inputs. These cues help them associate the visual information about objects with the audio signals possibly calling the properties of the object. Through several examples a young child is able to filter out the unnecessary information and better his/her association of the new objects with the ones previously learned.

In this paper we develop a novel approach in teaching robots the conceptual meanings of objects' properties. The robot is assumed to initially have only sensory capabilities i.e. visual (camera) and auditory (microphone) and no other knowledge about the world. Through several demonstrations carried out by the trainer the robot uses its visual sensory information to learn various object

properties from its field of view. As a parent does with a young child, the trainer shows an object to the robot while describing different properties of the object. For example, the trainer shows a green ball to the robot and speaks about its different properties, such as big, green, round and so on. This process is then repeated for different objects. From these examples the robot learns a representation of those properties, such as what it means to be big, green or round. When a novel object is presented to the robot, it tries to infer the properties of the object. Figure 1 gives an example of the training process. The first column shows objects and the second column lists the possible properties related to the objects. The first row shows one of the several examples of red objects while the second row shows one of the several examples of square objects. The last row shows an example of a novel object that was presented to the robot after the training process. The robot was able to infer that the novel object was red and square.




Objects	Trainer's comments about the properties of an object
	It is <i>Circular</i> It is <i>Big</i> It is <i>Red</i>
	It is <i>Rectangular</i> It is <i>Square</i> It is <i>Medium</i> It is <i>Blue</i>
A Novel Object	Robot's interpretation of the object properties
	It is <i>Square</i> It is <i>Red</i>

Figure 1: An example of objects and their properties.

Two techniques were investigated. In the first approach we assume no knowledge about the image features and their relevance to the conceptual information

relating to different objects. There are no pre-processing steps taken to extract information about object features and use them to train the system. The input to the system is the raw images taken by the camera and the image label, i.e. color, shape, etc. received from a microphone installed on the robot.

In the second approach we employ a pre-processing stage to extract various features from the image (e.g. color, number of corners). However, the robot does not know a priori the relevance between these features and the object properties. The goal of this method is to develop a technique for grouping the extracted features in order to infer an object's properties (e.g. rectangular, triangular etc.). For example the robot learns from the training examples that the object property shape is related to an extracted features i.e. number of corners. It eventually prunes its dictionary of concepts to maintain several meaningful key concepts, such as shapes, colors, orientation, etc.

Using either methods, the robot will be eventually capable of expanding its dictionary and better its ability to associate the object features it sees with the object properties it has already learned. Also, when the robot encounters new objects it will be able to infer that object's properties based on what it had learned during training process.

The rest of the paper is organized as follows. Section 2 discusses the related work. In section 3 we discuss in more detail the two proposed approaches. First a framework for Nonlinear Dimensionality Reduction, also called Isomap [1] is presented. Then the method based on feature extraction is presented and compared with the Isomap-based approach. Section 4 discusses the algorithms to address the aforementioned problem and results from both techniques are presented. Section 5 discusses the pros and cons of each technique and the future direction for the research.

2 Related work

Researchers from many different disciplines have focused on finding an effective and efficient way of learning objects. Learning processes vary depending on the domain they are applied to. Object learning (recognition) is considered to be a computer vision problem and relies mainly on computer vision techniques [2]. Learning of actions poses a different problem and has received more attention in the robotics community. Researchers have investigated various forms of social learning and interactive training techniques, e.g., learning by demonstration [3], tutelage [4], imitation-based learning [5], and clicker training [6]. This paper is an effort to find a novel approach that learns the meaning of an object's features as opposed

to the traditional approach that focuses on object recognition.

3 The tale of two approaches

In this section we present the background information about the two proposed techniques for learning of object properties and its application to our problem. The first subsection discusses the Isomap-based algorithm while the second subsection discusses the feature-based algorithm.

3.1 Isomap-based algorithm

In this algorithm the system does not have any a priori knowledge about the training images. The object of interest is assumed to have been segmented from the background and centered in the image using a nonparametric background subtraction algorithm [7]. An algorithm based on nonlinear dimensionality reductions is applied to the segmented images in the training data set.

The Isomap technique is used to reduce the dimensionality of a set of data such that the projection of the original data represents meaningful properties about the data set. There are several techniques in dimensionality reduction used in computer vision. Principal Component Analysis (PCA) [8] was used on images to recognize images of faces. PCA is considered a linear low-dimensional embedding mechanism which finds a low-dimensional embedding of data points which best preserves their variance. Another technique used in low-dimensional embedding is Multidimensional Scaling (MDS) [9]. It is used for exploring similarities and dissimilarities in data. Classical MDS finds an embedding which preserves the inter-point distance using Euclidean distance metrics. PCA and MDS are useful when the data contains only of a few and linear manifolds.

Many data sets may contain nonlinear structures that can not be captured by neither PCA nor MDS. In our case we want the system to learn about various objects, which may have various poses. This causes the data set to contain nonlinear manifolds. In order to account for the non-linear manifolds we use a non-linear low-dimensional data embedding called Isomap [1]. The Isomap technique uses geodesic distances of data points in the space which represents the non-linear geometry of the manifolds. For the points close to each other the input-space distances are good approximation to geodesic distances. For far away points the geodesic distance is approximated by adding the sequence of "short hops" between neighboring points [1].

The idea behind our first approach is to consider each object image as a high-dimensional data point. The raw images are fed to the Isomap-based algorithm, which generates a lower dimensional space. Objects of different sizes, shapes, colors, and orientations will be represented

on different partitions of this lower dimensional space when projected using Isomap. This allows us to find a re-partitioning of the data points projected into the lower dimensional space in such a way that the system automatically recognizes different properties of the objects.

3.2 Feature-based Algorithm

Like the Isomap-based method this algorithm does not have any knowledge about the images and the objects a priori. The object of interest is assumed to have been segmented from the background and centered in the image using the same non-parametric background subtraction algorithm as for the Isomap-based method. In this method the algorithm does not consider the images to be high-dimensional data points. Instead it employs a pre-processing technique to extract features from the images. However, the system does not consider any a priori information about what those features mean or the correlation between the extracted features and the properties of the object. During the learning process the system statistically correlates the object properties with the features shared by the most of the objects of the same type. Such process endows the system a relevance feature map. During the test stage, the relevance feature map is used to infer the object properties using the relevant features. Details about this algorithm are presented below.






Object					
No. of corners detected	3	6	0	5	12

Figure 2: A simple shape-related feature (No. of corners).

We used the OpenCV library to implement the feature extraction part of the system. We considered four main features:

- Color
- Size
- Shape
- Orientation

We consider that the system is presented with a single image of the object. The background is supposed to be black or the system needs an extra input of the masking image. The masking image needs to be with the object area as white and the background as black. The masking image can be extracted using the background segmentation if the sequence of images is used instead of a single image. The given image is processed using connected component analysis to find the biggest connected component for which the rest of processing (feature extraction) is performed.

For the purpose of the color extraction the average color of the object is determined. The color is presented in RGB (Red, Green, and Blue) space. However, HSI (Hue, Saturation, and Intensity) space might be more useful when conceptual learning is implemented. The hue and the saturation define the human perceived color, while the intensity defines the shade of the color. The transformation from RGB space to HSI space is straightforward. In case of the size, the horizontal and the vertical height of the object are determined. The area of the object is also determined. If the object is oblique (e.g. oblique ellipse), the major and the minor axis of the object are also determined. For determining the shape of the object, we consider the number of corners with the provision to add edge contour representation. Finding the dominant corners in an image is not an easy task as it looks to be. We start with finding the eigenvalues and eigenvectors for sufficiently large neighborhood (at least 5×5) over the entire image. Then, all the eigenvalues are sorted for a given neighborhood and the prominent corners are found with both of the eigenvalues high. The list of all the corners is created for the entire image and then searched to remove any corners that are very close to each other. Figure 2 shows an example of the images used and the number of corners detected by the corner detector. If the feature extraction algorithms used are robust enough, the features extracted using this method are simple enough to do direct classification.

We propose to use these basic features to formulate the learning strategy that can be used to understand concepts like color, shape and size. When enough data is collected demonstrating the specific concepts (e.g. small) then only the relevant basic features (or the relation between them) remain constant for the entire data set. Quantization of the feature space, the Gaussian mixture model for grouping the features and PCA for the reduction of dimension can help in learning these concepts. Even though the number of features considered here are small, it can be increased as the basic cues like texture and saliency are included. Different combination of the features present more features, e.g. to represent a square we need to consider the ratio between the width and height as this ratio is constant over the data set representing squares (not the actual width or height of the rectangle).

4 Results

In this section we present the results of the two proposed techniques for learning object properties. The first subsection discusses the results of the Isomap-based algorithm along with its application to SVM classifier to infer object properties of the novel objects. The second subsection discusses the results of the feature-based algorithm.

4.1 Isomap-based algorithm

In the following we present the results of the Isomap-based method to show its applicability to the learning mechanism we are addressing.

Figure 3 shows the results of the Isomap applied on a data set of 60 squares with 15 different sizes and 4 different shades of color. Figure 3-(a) shows a sample image of the database. Figure 3-(b) shows the geodesic distances between samples in the data set. The blue colors show smaller distances and the red colors correspond to the larger distances. The geodesic matrix is divided into 4 blocks rows and 4 block columns. From top-left to bottom-right, the first block (row 1, column 1) shows the geodesic distances of all the objects with the first shade. The second block (row 1, column 2) shows the geodesic distances of all objects with shade 1 from all objects with shade 2. As seen in the diagonal blocks, they have smaller geodesic distances since objects in those blocks share the same shades. Also the geodesic distances decrease as the size decreases. This is because there is less information for the smaller objects and therefore the total geodesic distances decrease.

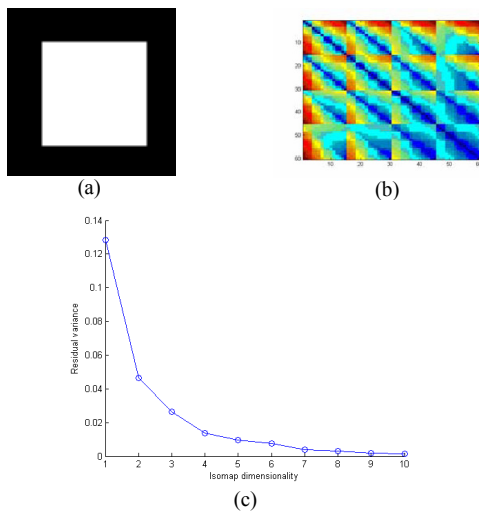


Figure 3: The Geodesic distance of 60 squares with 15 different sizes and 4 different shades and their Isomap space dimensions (a) A sample square image. (b) Geodesic distances. (c) Residual variances.

Finally the residual variances are shown in Figure 3-(c). As seen from the figure the variance of the data points decreases as the number of dimensions increase. This means that the Isomap method finds the dimensions which preserve the highest amount of variance of the data. By keeping the dimensions with highest residual variance we are keeping the most relevant information with the highest discriminatory information.

Figure 4 shows the Isomap space for a data set of 120 images containing 60 squares and 60 rectangles. The images contain objects of 15 different sizes and 4 different shades of gray. Figure 4-(a) shows the geodesic distance matrix for these images. As seen from the figure the matrix is divided into 4 major blocks each composed of 16 minor sub-blocks. The major block on the first row and the first column represents the geodesic distance of all the squares from each other. Since the matrix is symmetric the major block on the first row and second column contains the same geodesic distance as the block on the first column and second row. These blocks contain the geodesic distances of all the squares from the rectangles. The final major block contains the geodesic distances of all rectangles from the squares.

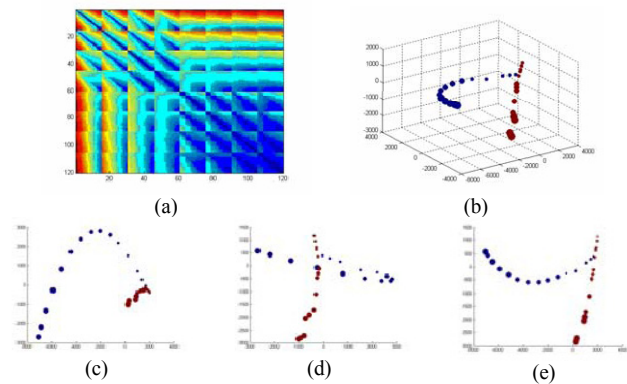


Figure 4: Isomap dimensionality reduction on a database of 120 images, 60 squares and 60 rectangles. (a) Geodesic distances (b) 3D low dimensional embedding (c)–(e) 3D projections.

The first observation from Figure 4-(a) is that as the size of the objects decreases the discriminatory factor of intra- and inter-class variations decreases, hence the blue colors for the latter entries in the sub-blocks and the diagonal entries. The second observation reveals that the rectangles in the data set on average were smaller than the squares (from the more blue colors) on the last major block.

From Figure 4-(b) we see the 3-dimensional embedding of the points. The red dots represent white rectangles of different sizes while the blue ones are the white squares. Also, the size of the dots is dependent on the size of the objects. Rectangles and squares of different colors are even more separated.

As seen there is a great potential in finding a linear discriminatory function which can easily separate the two classes. In order to investigate this option even further we project the two classes on a 2-D subspace of this three dimensional embedding space (Figure 4-(c)–Figure 4-(e)). As observed from these figures, the first dimension contains the variances of the squares and the second dimension encodes the rectangles. As we understand the last dimension of the embedding space represents the

prevalence of the shade feature. However, the latter observation needs to be more investigated and confirmed.

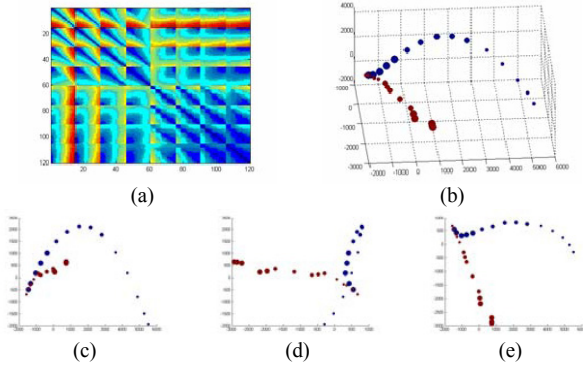


Figure 5: Isomap dimensionality reduction on a database of 120 images, 60 circles and 60 rectangles. (a) Geodesic distances (b) 3D low dimensional embedding (c)–(e) 2D projections.

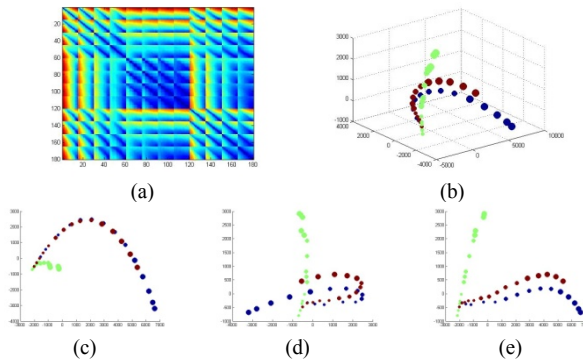


Figure 6: Isomap dimensionality reduction on a database of 180 images, 60 squares, 60 rectangles, 60 circles. (a) Geodesic distances (b) 3D low dimensional embedding (c)–(e) 2D projections.

Figure 5 shows the Isomap space for a data set of 120 images containing 60 circles and 60 rectangles. The images contain objects of 15 different sizes and 4 different shades of gray. Figure 5-(a) shows the geodesic distance matrix for these images. From Figure 5-(b) we see the 3-d dimensional embedding of the points. The red dots represent white rectangles of different sizes while the blue ones are the white circles. Rectangles and squares of different colors are even more separated. The projection of the two classes on a 2-D subspace of this three dimensional embedding space are shown in Figure 5-(c)–Figure 5-(e). As observed from these figures, the first dimension contains the variances of the circles and the second dimension encodes the rectangles.

Figure 6 shows the Isomap space for a data set of 180 images containing 60 squares, 60 rectangles, and 60 circles. The images contain objects of 15 different sizes (3 broad classes based on size: small, medium and large) and 4 different shades of gray. Figure 6-(a) shows the geodesic

distance matrix for these images. From Figure 6-(b) we see the 3-d dimensional embedding of the points. The green dots represent white rectangles of different sizes, the blue ones are the white circles and the red ones are white squares. The projection of the three classes on a 2-D subspace of this three dimensional embedding space are shown in Figure 6-(c)–Figure 6-(e). As observed from these figures, it is difficult to denote any particular dimension for a particular class (concept).

The robot is supposed to learn the conceptual information from the low dimensional data created by the Isomap. The Isomap was used to reduce the dimensionality to mere 10 dimensions. Therefore, each image is presented as a 10 dimensional feature vector. We considered three basic concepts that robot needs to understand, namely shape, size and color. The data is divided such that 80% of the data is used for training and the rest is reserved for testing. An SVM classifier [10] is trained and used to find how accurately it can classify shape, size and color. The Gaussian kernel was used to transform the data in higher dimension so that it is separable. Table 1 shows the results of SVM classifier on training and testing set for three different concepts viz. shape, size and color. The accuracy for shape and size recognition is excellent. However, the color recognition rate on test set is very small. This can be attributed to the number of samples used while training. The color information is separable, if only single shape is considered. However multiple shapes and sizes make this task difficult. The accuracy for color recognition can be increased if a hierarchical technique is used, where shape information is detected in the first step, followed by the color detection. We believe that the accuracy can also be increased by using direct technique like image processing for this purpose instead of Isomap.

Table 1: Accuracy of SVM classifier trained using the low dimensional features.

	Accuracy on Training set	Accuracy on Testing set
Shape	88.89%	83.33%
Size	95.83%	91.67%
Color	90.97%	50%

This shows a promising direction in terms of encoding the object features as dimensions of the embedding space. The current investigation demonstrates that each embedding dimension holds the key information about one property of an object, such as shape, color, etc. Therefore, the Isomap space can be immensely helpful in learning of currently known classes of properties and the novel ones. Secondly, the object property classification

(concept) using Isomaps becomes merely as finding of the dimension in the Isomap space with the largest variance.

We do not anticipate much deviation from the actual expected results by using real images, but these have not been tested so far. However, we anticipate a scaling problem with the Isomap, if we want to accommodate a large number of object features. Our expectation is that even if we could automatically infer which Isomap space dimension relates to which object property, by introducing more properties, such as colors, shapes, orientations, etc. we would require more dimensionality to encode these properties. Also, the problem of mapping a novel image to already known dimensions that were discovered during training phase of the Isomap poses a great challenge.

4.2 Feature-based algorithm

The feature based algorithm gives perfect results on the same dataset that was used for Isomap based approach. The reason for it is that feature extraction algorithm works the best on the synthetic dataset. Using the features to train SVM for feature based approach deemed unnecessary as features extracted (no. of corners, area, width, height and color) using feature extraction techniques directly reflect the properties of the objects (shape, size, and color). The feature based algorithm was an effort to show the reader how the same problem is handled by the researchers of computer vision. The problem with feature based algorithm is that it is not extendable to the more complex situations that include real world images and activities.

5 Conclusions

In this paper we investigated a developmental learning mechanism for robotic applications. Two approaches based on Isomap dimensionality reduction and feature-based learning mechanisms are investigated. Based on our experiments we have concluded that in the Isomap space, given enough training data and classes of categories, each dimension represents one or a combination of the object properties.

Our second approach uses a pre-processing technique to detect and extract pre-defined features from an object image. Although not completely unsupervised, due to the fact that the system is given the features to detect a priori, the algorithm does not take into account any relevance between the features and the object properties. The system learns as it is given more and more training samples the relevance factor and finally it is able to infer the objects and their properties from their features.

Our main investigation is concentrated in the Isomap space and finding of an automated inference mechanism for object learning. The Isomap technique was successful in

reducing dimensions to the meaningful features. The SVM classifier trained using these features was successful in determining the shape and the size of novel objects. The accuracy in determining color was low. However, the hierarchical classification technique can alleviate this problem. The current experiments used only synthetic images and all the images were used for the Isomap dimensionality reduction. The future work of this topic should consider using real images and try to tackle the problem of incremental Isomap technique which is able to reduce the dimensions of a novel image without repeating the entire process.

6 Acknowledgment

This work has been supported in part by NSF award IIS-0546876 and ONR awards N00014-06-1-0611 and N00014-09-1-1121.

7 References

- [1] J. Tenenbaum, V. de Silva, J. Langford, "A global geometric framework for nonlinear dimensionality reduction", *Science Magazine*, Vol. 290, pp. 2319-2323, 2000.
- [2] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning", *Proceedings of CVPR*, 2003.
- [3] M.N. Niculescu, M.J. Matari'c, "Natural methods for robot task learning: Instructive demonstrations, generalization and practice", *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*, Melbourne, Australia, July 2003.
- [4] C. Breazeal, G. Hoffman, A. Lockerd, "Teaching and working with robots as collaboration", *Proceedings of the AAMAS*, 2004.
- [5] S. Schaal, "Is imitation learning the route to humanoid robots?", *Trends in Cognitive Sciences* 3, 233-242, 1999.
- [6] B. Blumberg, M. Downie, Y. Ivanov, M. Berlin, M.P. Johnson, B. Tomlinson, "Integrated learning for interactive synthetic characters", *Proceedings of ACM SIGGRAPH 2002*, 2002.
- [7] A. Tavakkoli, M. Niculescu, M. Niculescu and G. Bebis, "Efficient Background Modeling through Incremental Support Vector Data Description", *Proceedings of ICPR*, 2008.
- [8] M. Turk, A. Pentland, "Face Recognition using Eigenfaces", *Proceedings of ICPR 1991*, pp. 586-591, 1991.
- [9] I. Borg, P. Groenen, "Multidimensional Scaling: Theory and Applications", Springer-Verlag, New York, 2005.
- [10] B. Schölkopf, A. Smola, R. Williamson, and P. L. Bartlett, "New support vector algorithms", *Neural Computation*, 12, pp. 1207-1245, 2000.