Content-Based Classification of CAD Models with Supervised Learning

Cheuk Yiu Ip and William C. Regli

Drexel University, ucip@cs.drexel.edu, regli@drexel.edu

ABSTRACT

This paper describes how to apply machine learning to adjust shape matching techniques to suit different classification of CAD models. Existing research based on either group technology or fixed modeling matching algorithms, impose a priori categorization schemes on engineering data or require significant human labeling of design data. This paper describes a general technique for "teaching" existing model comparison algorithms to adapt to different classifications that are relevant in many engineering applications. In this way, the core shape matching algorithms can be learned to adapt to wide variety of model classifications based on user input and training data. This allows for great flexibility in search and data mining of engineering data. Results are presented that show how to optimize different parameters to fit different categorization schema. The paper presents a comprehensive example and provides empirical results using a nearest neighbor classification on mechanical CAD models from the National Design Repository.

Keywords: CAD Database, Shape Matching, Machine Learning

1. INTRODUCTION

Indexing of parts and part families had been done with group technology (GT) coding in practice. Group technology was designed to facilitate process planning and cell-based manufacturing by imposing a classification schema on individual machined parts. These techniques were developed prior to inexpensive computer technology; hence they are not rigorously defined and are intended for human interpretation, rather than machine interpretation. Some of the early work on feature identification from solid models aimed to find patterns in model databases [14] or automate the GT coding [19,8,1] process [9,2,10,3]. The common aspect of all of these techniques is that they are all *post priori*: one runs their algorithm on model, and it produces the category or label for it.

Comparing 3D CAD Models

There are two basic types of approaches for matching and retrieval of 3D CAD data: (1) *feature-based* techniques and (2) *shape-based* techniques. The feature-based techniques [7,18], going back at least as far as the late 1970s [14], extract engineering features e.g. machining features, and form features, from a solid model of a mechanical part for use in database storage, and automated GT coding. Elinson et al. [6] used feature-based reasoning for retrieval of solid models for use in variant process planning. Cicirello and Regli [5] examined how to develop graph-based data structures and create heuristic similarity measures among artifacts; this work was extended in [4] to manufacturing feature-based similarity measurements. McWherter et al. [15] have integrated these ideas with database techniques to enable indexing and clustering of CAD models based on shape and engineering properties.

The shape-based techniques are more recent, owing to research contributions from computational geometry, computer vision, and computer graphics. A shape-based approach works as the representational "lowest common denominator": polygon mesh available from faceting solid models, in the form of VRML or STL. From the polygon mesh, different transformation invariant attributed can be extracted as the means of similarity among 3D models. Thompson et al. [20] examined reverse engineering of designs by generating surface and machining feature information from range data collected from machined parts. Hilaga et al. [11] present a method for matching 3D topological models using multi-resolution Reeb graphs. The method of Osada, Funkhouser et al [17] creates an abstraction of the 3D model as a probability distribution of samples from a shape function acting on the model. Kazhdan et al compares 3D models with spherical harmonics [13]. Novotni and Klein demonstrate the use of 3D zernike descriptors [16].

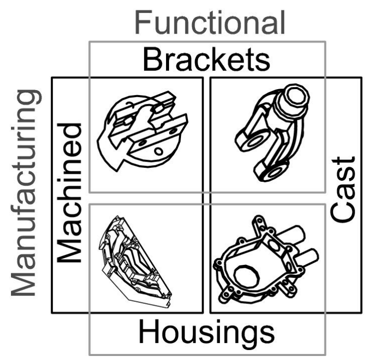


Fig. 1. Different Classification Schemas

The Research Challenge for Multi-Classification
Shape matching techniques are generally robust under model degradation, but it is a rigid technique and is a poor discriminator among model classes, because it usually emphasizes gross model shape, rather than the discriminatory features that are common in CAD/CAM data.

To complicate matters, in the CAD/CAM domain, engineering artifacts can have multiple classifications. For example, discrete machined parts can be classified in to different categories according to different classification criteria, such a functionality (e.g., brackets or fasteners), manufacturing cost and manufacturing process (e.g., casting, machining, forging, molding). Fig. 1 shows four CAD models under two different, but perfectly reasonably, classification schemas. The first classification is based on the manufacturing process, where parts are separated into either “3-axis machining” or “casting” processes. In machining, rotating cutting tools remove material based on swept volumes; these sweeps are limited to those on a 3-axis vertical machining center. The second, orthogonal classification, is based on mechanical function. Fig 1. also shows a decomposition into parts that function as “brackets” or as “housings”.

This paper describes how to apply machine learning to create automated classifiers for 3D CAD models. We describe a general technique for “teaching” existing model comparison algorithms to adapt to the different classifications that are relevant in many engineering applications. In this way, the core shape matching algorithms can be learned to adapt to wide variety of model classifications based on the user input and training data. Methods are presented that show how to optimize different parameters to fit different categorization schema. Lastly, a comprehensive example and is presented to provides empirical results using a nearest neighbor classification on mechanical CAD models from the National Design Repository (http://www.designrepository.org/).

2. LEARNING MULTIPLE CAD CLASSIFICATIONS

Research Goals
This research aims to develop automatic model classifiers by developing methods to train shape matching algorithms to discriminate among different model classes, such as those shown in Fig 1. Our approach optimizes performance for a particular classification schema and model comparison algorithm pair by adjusting the parameters of the model comparison algorithm. In this way, the shape matching technique is tuned to return shorter distances for models falling in the same category, but larger distance for models falling in different categories. Given a set of example CAD models and their corresponding categories, relevant features are weighted to automatically construct a model classifier. This work integrates traditional AI and machine learning with CAD and shape modeling. Specifically:

- By altering 3D model matching techniques, patterns or features of classification can be extracted from different perspectives to fit various classification schemas.
- Given different reasonable training examples, this approach can learn useful classification schemas to improve retrieval accuracy.

Hence, this research demonstrates a practical solution to a long-standing problem in CAD/CAM: automated part classification.

2.1. Applying Machine Learning to Classify CAD Models

Shape comparison algorithms provide different approaches to select invariant features from 3D models. These shape comparison algorithms transform a 3D model into a set of n comparable attributes, as a vector $< a_1, a_2, \cdots a_n >$

Switching among model comparison algorithms can focus classification schemas on different aspects, such as, topology, local geometry patterns, feature interactions, and gross shape. The flexibility of switching model comparison algorithms enables us to further optimize by matching comparison algorithm with classification schemas.

The aggregate distance between models in the same category should be relatively less than models falling into different categories. Given CAD models s_1, s_2, s_3, a category c_1 and the distance between s_1, s_2 as $D(s_1, s_2)$:

$$\forall_{s_1, s_2, s_3, c_1} : s_1, s_2 \in c_1 \wedge s_3 \notin c_1 \Rightarrow D(s_1, s_2) < D(s_1, s_3) \quad (1)$$

To improve the efficiency of learning classification schemas from training examples, distances produced by the model comparison algorithm, $D(s_1, s_2)$ should be adjusted to satisfy (1). Assuming $D(s_1, s_2)$ is produced by the gross difference of the two sets from n attributes representing s_1, s_2. The discriminatory power of each attribute can be studied; weights can be assigned to each attribute according to its significance in computing distance between example models and query models.

The distance in between a pair of models as the aggregate of some weighted distances among n attributes:

$$D(s_1, s_2) = \sum_{i=1}^{n} w_i \cdot D(s_{1_i}, s_{2_i})$$

$D(s_{1_i}, s_{2_i})$ represents the distance between attribute i of models s_1 and s_2, w_i represents the weight of the attribute. Evaluation of through machine learning techniques will be discussed in the next section.

3. TECHNICAL APPROACH

Shape distribution comparison functions [12, 17] is used as an example to illustrate the process of learning weights through a set of example models. Then the shape distributions and the learned parameters are used to classify incoming query models using nearest neighbor classification. The overall concept is shown in Fig. 2.

3.1. Primer on Shape Distribution Matching

A *shape distribution* can be viewed as a digital signature for a 3D model. Shape distribution techniques [17] are used with an enhancement for matching CAD models [12], to perform a statistical sampling of the shape properties of CAD models and use these samples to generate meaningful comparisons between the models.

Let s be a CAD model, let $T = \{t_1, t_2 \cdots t_k\}$ be a set of triangular facets that approximate the topology of s. The facets of T could be produced by existing mesh generation algorithms, Stereolithography exporters or VRML exporters. The facets in T could also be from an active data acquisition system working off of actual models, as in [20].

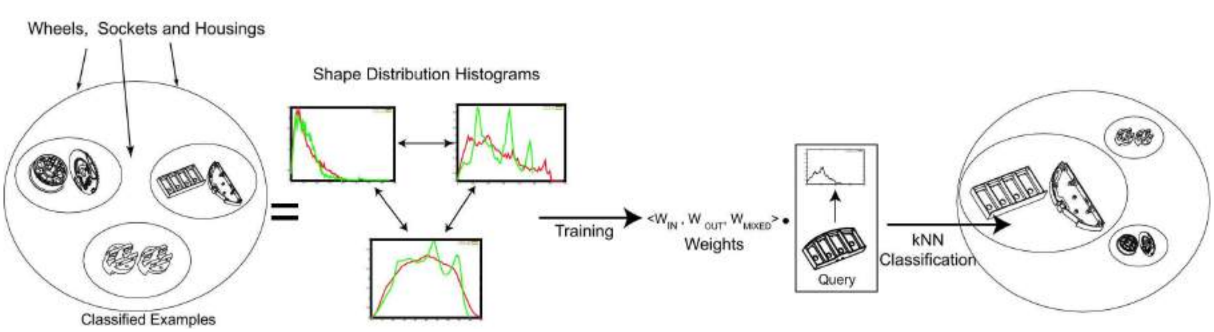


Fig. 2. Technical Approach Overview

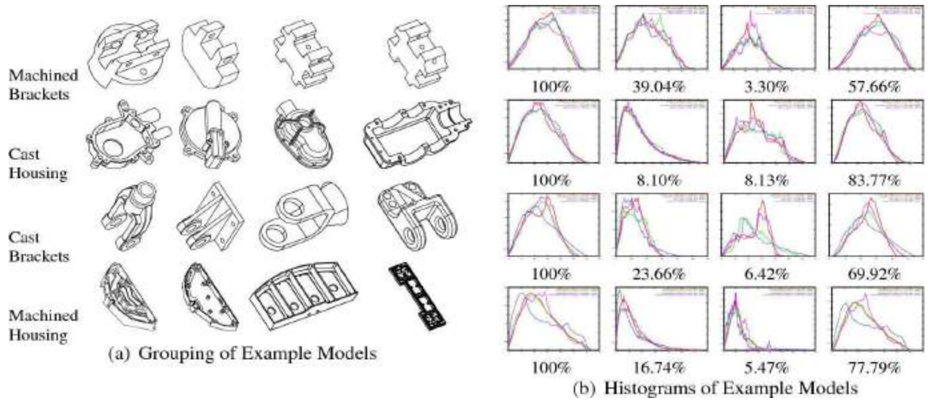


Fig 3. Example Classified Models and Shape

Matching two models with shape distributions requires:

1. **Selecting a shape function.** Many different functions were tested by [17] and the $D2$ shape function (measuring the distance between two random points on the surface of a model) was shown to be the most consistent.
2. **Sampling of random points.** Generate a sufficiently large number of random sample point pairs on surface of model s.
3. **Classification of point-pair distances.** [12] In computing the $D2$, distances are classified based on their interaction with the local geometry and topology. Specifically, there are three kinds of interactions:
    - IN distances: The line connecting the two points lies completely inside the model.
    - OUT distances: The line connecting the two points lies completely outside the model.
    - MIXED distances: The line connecting the two points passes both inside and outside of the model.
4. **Calculate shape distribution histograms.** Construct histograms associated with *IN*, *OUT* and *MIXED* $D2$ shape distribution function.
5. **Compare shape distribution histograms.** Using well-known curve matching techniques (e.g. Minkowski L_N distance).

3.1.1. Comparing Shape Distributions

Shape distributions histograms are compared to produce dissimilarity measures. The dissimilarity in between models is represented by a per bin L_1 norm Minkowski distance in between their corresponding shape distribution histograms, computed using across each of the j histogram bins as:

$$L(h_1, h_2) = \frac{\sum_{i=0}^{n} | h_{1_i} - h_{2_i} |}{j}$$

This is done for each of the *IN*, *OUT*, and *MIXED* histograms.

Example
16 CAD models classified in four categories along with their histograms are shown in Fig. 3.

3.2. Learning Optimal Weights

Each histogram carries different degrees of significance in distance computation $< w_{IN}, w_{OUT}, w_{MIXED} >$. Distance function should produce smaller distance for parts sharing the same category, and larger distance for parts falling into different categories, such that, the weighted distance of histograms, $w_{IN} \cdot IN + w_{OUT} \cdot OUT + w_{MIXED} \cdot MIXED$ minimizes the aggregate distance within a category and maximize the aggregate distance across different categories.

3.2.1. Training Process

Shape distribution histograms of all training examples are being compared to find out the frequencies of *IN*, *OUT* or *MIXED* distance being selected to be representative distance. Provided a set of training examples models, the training of weights for *IN*, *OUT* or *MIXED* histograms can be carried out as follow:

1. Profile the example models and construct *IN*, *OUT* and *MIXED* histograms as signatures of each model.
2. Compute the L_1 Minkowski distances in between all pairs of example models.
3. Pick the appropriate *IN*, *OUT* or *MIXED* distance for each pair of models, according to the corresponding categories, as the representative distance.
4. Normalize the frequencies of *IN*, *OUT* or *MIXED* distance being selected as representative distances to be the weights.

Three distances, *IN*, *OUT* and *MIXED,* are calculated for each pair of models. From each triplet of distances, only one suitable representative distance is selected for weights computation.

- Select the shortest distance among *IN*, *OUT* and *MIXED* when two models fall into the same category.
- Select the longest distance among *IN*, *OUT* and *MIXED* when two models fall into the different categories.

The frequencies of the *IN*, *OUT*, and *MIXED* distances are selected as representative distances to reflect that the chance of *IN*, *OUT*, and *MIXED* distance being the appropriate distance for aggregate distance computation. The weights triplet $< w_{IN}, w_{OUT}, w_{MIXED} >$ is derived as:

$$w_i = \frac{\# i}{\# IN + \# OUT + \# MIXED}, i \in \{IN, OUT, MIXED\}$$

Different weight triplets are produced to fit each category specifically. When computing distance in between a query model and an example model, the weight triplets that corresponds to the example model's category will be used to scale the distance of *IN*, *OUT* and *MIXED* histograms.

Example

Training example models presented in Fig. 3 are grouped into four different categories: Prismatic machined Brackets, Cast-then-machined Housings, Cast-then-machined Brackets and Prismatic machined Housings. To demonstrate our training technique can learn different weighting schemas from examples, four groups of models are classified into two different classes.

- Functional — Brackets or Housings (Tab. 1)
- Manufacturing — Prismatic Machined or Cast-then-machined (Tab. 2)

Pairwise distances between all models are computed in Tab. 1(a) (Functional classification) and Tab 2(a). (Manufacturing classification) to obtain two different sets of weights; boxed distances in the tables denotes representative distances.

Tab. 1(b) and Tab 2(b) summarize different weighting schemes captured from the example models. For example, in functional classification, 47 *IN* distances, 23 *OUT* distances and 46 *MIXED* distances are selected as representative distances. By normalizing the proportion of representative distances, a weighting scheme for functional classification could be 39% for *IN*, 23% for *OUT* and 38% *MIXED*. Weighting schemes learned from this process will be used as input during nearest neighbors' classification process.

3.3. Nearest Neighbor Classification

The classification of the query model s_q is computed and returned after this process. Considering the categories of the k nearest example models, the classification of the query model s_q is determined by a locally weighted function provided to the nearest neighbor classification. Here are two simple variations of the function:

(a) Pairwise Distances Between Example Models

	Brackets								Housings							
I	47.82	52.43	55.33	46.27	17.26	33.52	45.46	41.12	10.64	6.64	11.58	24.95	9.99	8.23	11.47	0
O	32.11	72.11	28.52	26.83	38.96	42.42	81.77	72.01	26.37	25.47	18.44	31.17	36.31	37.55	28.43	0
M	68.55	52.45	54.27	58.97	13.52	47.23	76.53	55.73	24.61	21.39	23.38	77.99	19.43	26.19	21.82	0
I	37.95	42.71	45.09	36.00	9.10	28.81	37.45	31.64	15.90	8.84	15.10	33.08	12.60	6.84	0	11.47
O	55.06	81.02	40.68	37.56	33.64	56.39	86.70	73.89	24.25	31.82	31.90	29.91	23.87	34.82	0	28.43
M	50.37	39.55	34.79	39.24	19.34	29.22	55.22	36.61	31.42	29.11	17.42	91.83	19.46	18.93	0	21.82
I	40.89	45.15	48.16	38.83	11.62	29.04	39.09	34.76	13.82	7.08	13.09	29.33	9.41	0	6.84	8.23
O	58.75	97.85	40.94	41.11	47.42	63.78	89.24	80.88	35.80	41.39	27.68	30.91	27.90	0	34.82	37.55
M	62.54	35.60	41.50	45.32	25.16	39.25	62.37	52.13	34.92	31.66	26.60	92.22	15.86	0	18.93	26.19
I	42.83	46.66	51.08	41.01	16.39	25.91	42.01	36.42	12.78	10.19	10.67	29.99	0	9.41	12.60	9.99
O	55.18	84.61	38.01	37.64	42.81	59.13	79.51	67.23	27.00	37.83	28.40	31.19	0	27.90	23.87	36.31
M	63.45	45.24	47.22	49.96	16.88	44.22	62.36	51.37	23.00	19.94	17.84	71.47	0	15.86	19.46	19.43
I	60.61	64.92	66.49	59.74	38.77	46.92	57.21	56.08	27.59	21.98	31.08	0	29.99	29.33	33.08	24.95
O	34.43	46.59	34.13	33.60	25.66	33.40	46.77	44.74	24.92	24.99	30.75	0	31.19	30.91	29.91	31.17
M	141.08	124.18	124.43	128.08	81.16	119.95	134.66	122.86	60.79	65.28	79.60	0	71.47	92.22	91.83	77.99
I	41.55	47.09	50.86	40.65	18.06	25.73	40.51	36.81	7.49	12.36	0	31.08	10.67	13.09	15.10	11.58
O	18.42	47.68	23.02	25.11	35.90	29.05	51.60	46.08	30.22	30.90	0	30.75	28.40	27.68	31.90	18.44
M	64.58	53.05	47.53	51.26	24.65	42.39	63.95	47.54	25.59	24.55	0	79.60	17.84	26.60	17.42	23.38
I	35.53	38.73	40.78	34.03	13.83	26.44	34.44	30.86	12.14	0	12.36	21.98	10.19	7.08	8.84	6.64
O	33.40	64.46	33.91	33.05	22.76	31.37	72.93	68.05	19.88	0	30.90	24.99	37.83	41.39	31.82	25.47
M	75.67	57.29	61.16	64.72	19.04	52.91	79.43	63.91	10.23	0	24.55	65.28	19.94	31.66	29.11	21.39
I	42.83	47.89	50.86	41.71	20.26	28.61	40.88	39.25	0	12.14	7.49	27.59	12.78	13.82	15.90	10.64
O	45.37	73.12	34.06	34.21	24.62	42.53	77.06	68.39	0	19.88	30.22	24.92	27.00	35.80	24.25	26.37
M	78.11	60.22	63.20	66.84	22.64	56.38	81.23	65.83	0	10.23	25.59	60.79	23.00	34.92	31.42	24.61
I	13.02	24.01	22.70	21.76	24.55	32.37	20.38	0	39.25	30.86	36.81	56.08	36.42	34.76	31.64	41.12
O	79.34	48.71	52.16	48.02	86.07	88.90	39.77	0	68.39	68.05	46.08	44.74	67.23	80.88	73.89	72.01
M	29.61	53.20	28.88	29.65	55.84	24.08	40.17	0	65.83	63.91	47.54	122.86	51.37	52.13	36.61	55.73
I	19.92	26.74	26.94	33.01	31.48	37.65	0	20.38	40.88	34.44	40.51	57.21	42.01	39.09	37.45	45.46
O	86.74	50.79	60.32	55.26	95.47	93.59	0	39.77	77.06	72.93	51.60	46.77	79.51	89.24	86.70	81.77
M	32.22	63.12	33.10	29.39	75.86	45.53	0	40.17	81.23	79.43	63.95	134.66	62.36	62.37	55.22	76.53
I	35.28	37.17	45.70	27.68	29.52	0	37.65	32.37	28.61	26.44	25.73	46.92	25.91	29.04	28.81	33.52
O	31.22	79.13	43.10	40.16	45.52	0	93.59	88.90	42.53	31.37	29.05	33.40	59.13	63.78	56.39	42.42
M	27.38	32.62	22.25	23.97	43.33	0	45.53	24.08	56.38	52.91	42.39	119.95	44.22	39.25	29.22	47.23
I	29.12	36.48	36.85	28.23	0	29.52	31.48	24.55	20.26	13.83	18.06	38.77	16.39	11.62	9.10	17.26
O	50.02	89.59	51.90	47.47	0	45.52	95.47	86.07	24.62	22.76	35.90	25.66	42.81	47.42	33.64	38.96
M	65.49	47.45	52.17	56.99	0	43.33	75.86	55.84	22.64	19.04	24.65	81.16	16.88	25.16	19.34	13.52
I	17.92	30.37	28.94	0	28.23	27.68	33.01	21.76	41.71	34.03	40.65	59.74	41.01	38.83	36.00	46.27
O	30.62	45.93	14.97	0	47.47	40.16	55.26	48.02	34.21	33.05	25.11	33.60	37.64	41.11	37.56	26.83
M	22.12	34.60	9.90	0	56.99	23.97	29.39	29.65	66.84	64.72	51.26	128.08	49.96	45.32	39.24	58.97
I	20.49	28.30	0	28.94	36.85	45.70	26.94	22.70	50.86	40.78	50.86	66.49	51.08	48.16	45.09	55.33
O	31.98	54.62	0	14.97	51.90	43.10	60.32	52.16	34.06	33.91	23.02	34.13	38.01	40.94	40.68	28.52
M	26.52	32.22	0	9.90	52.17	22.25	33.10	28.88	63.20	61.16	47.53	124.43	47.22	41.50	34.79	54.27
I	22.34	0	28.30	30.37	36.48	37.17	26.74	24.01	47.89	38.73	47.09	64.92	46.66	45.15	42.71	52.43
O	70.20	0	54.62	45.93	89.59	79.13	50.79	48.71	73.12	64.46	47.68	46.59	84.61	97.85	81.02	72.11
M	46.22	0	32.22	34.60	47.45	32.62	63.12	53.20	60.22	57.29	53.05	124.18	45.24	35.60	39.55	52.45
I	0	22.34	20.49	17.92	29.12	35.28	19.92	13.02	42.83	35.53	41.55	60.61	42.83	40.89	37.95	47.82
O	0	70.20	31.98	30.62	50.02	31.22	86.74	79.34	45.37	33.40	18.42	34.43	55.18	58.75	55.06	32.11
M	0	46.22	26.52	22.12	65.49	27.38	32.22	29.61	78.11	75.67	64.58	141.08	63.45	62.54	50.37	68.55

(b) Derived weights for Functional Classification

Category	*IN*	*OUT*	*MIXED*
Brackets	39% (47)	23% (27)	38% (46)
Housings	44% (53)	28% (33)	28% (34)

Tab. 1 Functional Classification

(a) Pairwise Distances Between Example Models

	Machined								Molded							
I	47.82	52.43	55.33	46.27	10.64	6.64	11.58	24.95	17.26	33.52	45.46	41.12	9.99	8.23	11.47	0
O	32.11	72.11	28.52	26.83	26.37	25.47	18.44	31.17	38.96	42.42	81.77	72.01	36.31	37.55	28.43	0
M	68.55	52.45	54.27	58.97	24.61	21.39	23.38	77.99	13.52	47.23	76.53	55.73	19.43	26.19	21.82	0
I	37.95	42.71	45.09	36.00	15.90	8.84	15.10	33.08	9.10	28.81	37.45	31.64	12.60	6.84	0	11.47
O	55.06	81.02	40.68	37.56	24.25	31.82	31.90	29.91	33.64	56.39	86.70	73.89	23.87	34.82	0	28.43
M	50.37	39.55	34.79	39.24	31.42	29.11	17.42	91.83	19.34	29.22	55.22	36.61	19.46	18.93	0	21.82
I	40.89	45.15	48.16	38.83	13.82	7.08	13.09	29.33	11.62	29.04	39.09	34.76	9.41	0	6.84	8.23
O	58.75	97.85	40.94	41.11	35.80	41.39	27.68	30.91	47.42	63.78	89.24	80.88	27.90	0	34.82	37.55
M	62.54	35.60	41.50	45.32	34.92	31.66	26.60	92.22	25.16	39.25	62.37	52.13	15.86	0	18.93	26.19
I	42.83	46.66	51.08	41.01	12.78	10.19	10.67	29.99	16.39	25.91	42.01	36.42	0	9.41	12.60	9.99
O	55.18	84.61	38.01	37.64	27.00	37.83	28.40	31.19	42.81	59.13	79.51	67.23	0	27.90	23.87	36.31
M	63.45	45.24	47.22	49.96	23.00	19.94	17.84	71.47	16.88	44.22	62.36	51.37	0	15.86	19.46	19.43
I	13.02	24.01	22.70	21.76	39.25	30.86	36.81	56.08	24.55	32.37	20.38	0	36.42	34.76	31.64	41.12
O	79.34	48.71	52.16	48.02	68.39	68.05	46.08	44.74	86.07	88.90	39.77	0	67.23	80.88	73.89	72.01
M	29.61	53.20	28.88	29.65	65.83	63.91	47.54	122.86	55.84	24.08	40.17	0	51.37	52.13	36.61	55.73
I	19.92	26.74	26.94	33.01	40.88	34.44	40.51	57.21	31.48	37.65	0	20.38	42.01	39.09	37.45	45.46
O	86.74	50.79	60.32	55.26	77.06	72.93	51.60	46.77	95.47	93.59	0	39.77	79.51	89.24	86.70	81.77
M	32.22	63.12	33.10	29.39	81.23	79.43	63.95	134.66	75.86	45.53	0	40.17	62.36	62.37	55.22	76.53
I	35.28	37.17	45.70	27.68	28.61	26.44	25.73	46.92	29.52	0	37.65	32.37	25.91	29.04	28.81	33.52
O	31.22	79.13	43.10	40.16	42.53	31.37	29.05	33.40	45.52	0	93.59	88.90	59.13	63.78	56.39	42.42
M	27.38	32.62	22.25	23.97	56.38	52.91	42.39	119.95	43.33	0	45.53	24.08	44.22	39.25	29.22	47.23
I	29.12	36.48	36.85	28.23	20.26	13.83	18.06	38.77	0	29.52	31.48	24.55	16.39	11.62	9.10	17.26
O	50.02	89.59	51.90	47.47	24.62	22.76	35.90	25.66	0	45.52	95.47	86.07	42.81	47.42	33.64	38.96
M	65.49	47.45	52.17	56.99	22.64	19.04	24.65	81.16	0	43.33	75.86	55.84	16.88	25.16	19.34	13.52
I	60.61	64.92	66.49	59.74	27.59	21.98	31.08	0	38.77	46.92	57.21	56.08	29.99	29.33	33.08	24.95
O	34.43	46.59	34.13	33.60	24.92	24.99	30.75	0	25.66	33.40	46.77	44.74	31.19	30.91	29.91	31.17
M	141.08	124.18	124.43	128.08	60.79	65.28	79.60	0	81.16	119.95	134.66	122.86	71.47	92.22	91.83	77.99
I	41.55	47.09	50.86	40.65	7.49	12.36	0	31.08	18.06	25.73	40.51	36.81	10.67	13.09	15.10	11.58
O	18.42	47.68	23.02	25.11	30.22	30.90	0	30.75	35.90	29.05	51.60	46.08	28.40	27.68	31.90	18.44
M	64.58	53.05	47.53	51.26	25.59	24.55	0	79.60	24.65	42.39	63.95	47.54	17.84	26.60	17.42	23.38
I	35.53	38.73	40.78	34.03	12.14	0	12.36	21.98	13.83	26.44	34.44	30.86	10.19	7.08	8.84	6.64
O	33.40	64.46	33.91	33.05	19.88	0	30.90	24.99	22.76	31.37	72.93	68.05	37.83	41.39	31.82	25.47
M	75.67	57.29	61.16	64.72	10.23	0	24.55	65.28	19.04	52.91	79.43	63.91	19.94	31.66	29.11	21.39
I	42.83	47.89	50.86	41.71	0	12.14	7.49	27.59	20.26	28.61	40.88	39.25	12.78	13.82	15.90	10.64
O	45.37	73.12	34.06	34.21	0	19.88	30.22	24.92	24.62	42.53	77.06	68.39	27.00	35.80	24.25	26.37
M	78.11	60.22	63.20	66.84	0	10.23	25.59	60.79	22.64	56.38	81.23	65.83	23.00	34.92	31.42	24.61
I	17.92	30.37	28.94	0	41.71	34.03	40.65	59.74	28.23	27.68	33.01	21.76	41.01	38.83	36.00	46.27
O	30.62	45.93	14.97	0	34.21	33.05	25.11	33.60	47.47	40.16	55.26	48.02	37.64	41.11	37.56	26.83
M	22.12	34.60	9.90	0	66.84	64.72	51.26	128.08	56.99	23.97	29.39	29.65	49.96	45.32	39.24	58.97
I	20.49	28.30	0	28.94	50.86	40.78	50.86	66.49	36.85	45.70	26.94	22.70	51.08	48.16	45.09	55.33
O	31.98	54.62	0	14.97	34.06	33.91	23.02	34.13	51.90	43.10	60.32	52.16	38.01	40.94	40.68	28.52
M	26.52	32.22	0	9.90	63.20	61.16	47.53	124.43	52.17	22.25	33.10	28.88	47.22	41.50	34.79	54.27
I	22.34	0	28.30	30.37	47.89	38.73	47.09	64.92	36.48	37.17	26.74	24.01	46.66	45.15	42.71	52.43
O	70.20	0	54.62	45.93	73.12	64.46	47.68	46.59	89.59	79.13	50.79	48.71	84.61	97.85	81.02	72.11
M	46.22	0	32.22	34.60	60.22	57.29	53.05	124.18	47.45	32.62	63.12	53.20	45.24	35.60	39.55	52.45
I	0	22.34	20.49	17.92	42.83	35.53	41.55	60.61	29.12	35.28	19.92	13.02	42.83	40.89	37.95	47.82
O	0	70.20	31.98	30.62	45.37	33.40	18.42	34.43	50.02	31.22	86.74	79.34	55.18	58.75	55.06	32.11
M	0	46.22	26.52	22.12	78.11	75.67	64.58	141.08	65.49	27.38	32.22	29.61	63.45	62.54	50.37	68.55

(b) Derived weights for Manufacturing Classification

Category	*IN*	*OUT*	*MIXED*
Machined	23% (30)	40% (52)	37% (48)
Cast	48% (57)	24% (29)	28% (34)

Tab. 2 Manufacturing Classification

- The *Majority* method simply returns the majority of categories of the k nearest training example neighbors as the classification. All k example neighbors *vote* for their classification. The classification with the highest votes will be returned as the classification of s_q.
- *Gaussian Kernel Regression* assigns weights to the k nearest example neighbors, s_i, according to a Gaussian kernel function. $D(s_i, s_q)$ is the distance in between, example model s_i and query model s_q, with a standard deviation σ.

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{D(s_i,s_q)^2}{2\sigma}}$$

The category with the highest *accumulated* weight is returned as the classification of s_q.

Example

Using the example models and weighting schema captured in the last section, the classification process is illustrated by classifying a query model, `base2` (Fig. 4). `Base2` is a bracket and manufactured by prismatic machining processes.

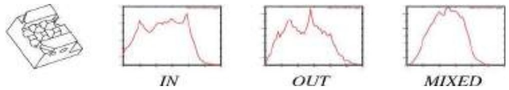


Fig. 4. Query Model `base2` and Histograms

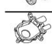
The classification process with both weighted functions on functional and manufacturing classifications:

1. Computes the distance and apply an appropriate weighting scheme for each category to obtain a weighted distance between example models and the query model. Tab. 3 and Tab. 4 shows the resulting distances and classifications between query models and example models, under both functional and manufacturing classifications.
2. Use the *majority* or the *Gaussian Kernel Regression* method to classify the query model
   a. Using a *majority* method to classify the query model, k =5 is used in this example, Five of the closest training models are picked to "vote" for the classification of the query model. The five closest weighted distances in Tab. 3 (functional classification) and Tab. 4 (manufacturing classification) are boxed. Under the functional classification weighting schema (Tab. 3), query model `base2` is classified as a bracket, as four out of five of its closest neighbors come from the Bracket category. Under the manufacturing classification weighting schema, query model `base2` is classified as a Prismatic Machined part, as three out of five of its closest neighbors come from the prismatic machined parts category.
   b. Using the *Gaussian Regression* method to classify the query model, the resulting Gaussian weight of each example is showed in Tab. 3 and Tab. 4, the total cumulative Gaussian weights are present in bold face, and the categories with higher weights are boxed. The classification result is the same as using the *majority* method, the query model is classified into Brackets and Prismatic Machined categories. Note that this regression method is biased towards the closest model, which may generate a weight that dominates the others.

This example demonstrates that the different weights learned in the previous steps can be use to effectively to classify query models under different classifications schemas, along with the same set of training models.

	IN	*OUT*	*MIXED*	Weighted	Gaussian Regression	
	34.81	48.14	37.43	$\boxed{38.81}$		7.75E-36
	35.26	76.33	22.93	39.77		1.58E-37
	34.06	28.03	16.46	$\boxed{25.96}$		3.40E-17
	32.42	32.14	21.73	$\boxed{28.26}$		5.46E-20
	41.84	53.75	48.41	47.04		1.21E-51
	34.82	56.89	28.73	$\boxed{37.45}$		1.64E-33
	37.49	65.69	40.56	45.01		1.81E-47
	32.47	56.84	37.58	39.91	8.98E-38	$\boxed{\textbf{Total:3.41E-17}}$
	39.92	42.92	55.99	45.30		4.78E-48
	47.75	37.08	58.38	47.83		2.55E-53
	46.37	24.55	43.86	$\boxed{39.66}$		2.54E-37
	61.18	34.00	119.29	70.17		2.44E-112
	48.51	30.00	41.54	41.44		1.45E-40
	47.90	30.18	34.12	$\boxed{39.12}$		2.24E-36
	46.28	41.00	30.63	47.05		1.22E-38
	52.99	36.75	51.07	47.98	1.20E-53	**Total:2.51E-36**

Tab. 3 Distances between `base2` and Training Models (Functional)

	IN	*OUT*	*MIXED*	Weighted	Gaussian Regression	
	34.81	48.14	37.43	41.11		6.05E-40
	35.26	76.33	22.93	47.14		7.54E-52
	34.06	28.03	16.46	$\boxed{25.15}$		2.85E-16
	32.42	32.14	21.73	$\boxed{28.36}$		4.05E-20
	39.92	42.92	55.99	47.05		1.12E-51
	47.75	37.08	58.38	47.41		2.01E-52
	46.37	24.55	43.86	$\boxed{36.72}$		2.74E-32
	61.18	34.00	119.29	71.76	2.10E-117	$\boxed{\textbf{Total:2.85-E16}}$
	41.84	53.75	48.41	46.58		1.10E-50
	34.82	56.89	28.73	$\boxed{38.43}$		3.60E-35
	37.49	65.69	40.56	45.17		8.49E-48
	32.47	56.84	37.58	39.81		1.38E-37
	48.51	30.00	41.54	42.06		1.02E-41
	47.90	30.18	34.12	$\boxed{39.71}$		2.03E-37
	46.28	41.00	30.63	40.57		5.86E-39
	52.99	36.75	51.07	48.52.98	8.12E-55	**Total:3.63E-35**

Tab. 4 Distances between `base2` and Training Models (Manufacturing)

4. EMPIRICAL RESULTS AND DISCUSSION

To validate this approach, it was applied to learn and classify an expanded data set of 100 CAD models, according to manufacturing properties. This method has been implemented in Java/Perl, and executed on Linux platforms.

Experiments were conducted using a subset of mechanical part data from the National Design Repository. The model datasets are available at http://www.designrepository.org/datasets/.

The datasets was initially classified by hand into (1) prismatic machined parts and (2) parts that are first cast and then have their finishing features machined. The engineering rationale in this classification is that parts that are exclusively machined are usually high-precision parts, or parts made in small batches (i.e., for custom jobs). Cast-then-machined parts are typically from larger production runs and generally have much looser tolerance considerations for the non-machined surfaces of the object. In this case the investment of physical plant is larger, as is the manufacturing production plan (i.e., one needs to machine a mold for casting).

Training examples were randomly selected based on this classification. The objective was to see if the system could learn weights from the training examples and then classify the non-training examples in a manner consistent with the human classification. The experiment was repeatedly performed to confirm the robustness of our approach. The results are summarized in Tab. 5.

	Correctness
Highest	76%
Average	63%

Tab. 5. Classification Statistics

In these experiments the highest classification correctness reached 76%, with an average classification correctness of 63%. While 63% might not sound like a particularly good score, it needs to be noted that:

- There are no other fully automated, customizable, part classification schemes for 3D CAD data.
- As with any automated classification system, performance depends on the quality of training examples and the make up of the overall data set. In the experiments, data was sampled from the National Design Repository. While the CAD models selected are all real engineering artifacts, they exhibit considerable heterogeneity and variability within classes. The technique is expected to perform better in a more realistic setting where the datasets would be much larger and more structurally homogeneous within class. A larger number of training examples can also be provided to increase accuracy.
- Better training sets can be identified. Classification varies with training dataset; hence some training models would lead to a better performance. Instead of providing random examples as in our experiments, hand picking examples by experts can possibly improve the accuracy of classification.

5. CONCLUSIONS AND FUTURE WORK

This paper described a new approach to automate the classification of CAD models with machine learning techniques. The paper shows how existing work in 3D matching can be extended to enable data mining in the challenging new domain of CAD databases.

The contribution of this research is applying machine learning to shape matching to categorize CAD models, both old archived components and brand-new parts, with any existing classification schema. At a minimum, this approach represents a method for augmenting traditional group technology coding schemes by providing a completely digital process for storage and comparison of engineering models. With refinement, eventually users will be able to *train* a retrieval system to recognize arbitrary categories based on shape, manufacturing or other properties. With long-term research, we believe that the techniques in this paper could replace GT schemes with flexible matching systems adaptable to diverse engineering disciplines.

Acknowledgments

This work was supported in part by National Science Foundation (NSF) CAREER Award CISE/IIS-9733545 and Office of Naval Research (ONR) Grant N00014-01-1-0618. Additional support has been provided by Honeywell FM&T.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the other supporting government and corporate organizations.

6. REFERENCES

[1] Ames. A. L. Production ready feature recognition based automatic group technology part coding. In Symposium on Solid Modeling Foundations and CAD/CAM Applications, 1991, pp 161-169.

[2] Anderson D. C. and Henderson M. R.. Computer recognition and extraction of form features: A CAD/CAM link. Computers in Industry, Vol 5, 1984, pp 329-339

[3] Anderson D. C., M. R. Henderson M. R., and Staley S. M.. Using syntactic pattern recognition to extract feature information from a solid geometric data base. Computers in Mechanical Engineering, 1983.

[4] Cicirello V. and W. Regli W. C.. Machining feature-based comparisons of mechanical parts. In Shape Modeling and Applications, 2001, pp 176-187.

[5] Cicirello V. and Regli W. C.. Resolving non-uniqueness in design feature histories. In Fifth Symposium on Solid Modeling and Applications, 1999, pp 76-84.

[6] Elinson A., Nau D. S., and Regli W. C.. Feature-based similarity assessment of solid models. In Fourth Symposium on Solid Modeling and Applications, 1997, pp 297-310.

[7] Han J.-H., Regli W. C., and Pratt M. J.. Algorithms for feature recognition from solid models: A status report. IEEE Transactions on Robotics and Automation, Vol. 16, 2000, pp 782-796.

[8] Harhalakis G., Lin C.-P., Mark L., and Muro-Medrano P.. Implementation of rule-based information systems for integrated manufacturing. IEEE Transactions on Knowledge and Data Engineering, Vol. 6, 1994, pp 892-908.

[9] Henderson M. and Musti S.. Automated group technology part coding from a three-dimensional cad database. Journal of Engineering for Industry, Vol. 110, 1998 pp 278-287.

[10] Henderson M. R.. Extraction of Feature Information from Three-Dimensional CAD Data. PhD thesis, Purdue University, West Lafayette, IN, USA, 1984.

[11] Hilaga M., Shinagawa Y., Kohmura T., and Kunii T. L.. Topology matching for fully automatic similarity estimation of 3d shapes. In SIGGRAPH, 2001, pp 203-212.

[12] Ip C. Y., Lapadat D., Sieger L., and Regli W. C.. Using shape distributions to compare solid models. In Seventh Symposium on Solid Modeling and Applications, 2002, pp 273-280

[13] Kazhdan M., Funkhouser T., and Rusinkiewicz S.. Rotation invariant spherical harmonic representation of 3d shape descriptors. In Eurographics/ACM SIGGRAPH symposium on Geometry processing, 2003, pp 156-164.

[14] Kyprianou L. K.. Shape Classification in Computer Aided Design. PhD thesis, Christ College, University of Cambridge, Cambridge, United Kingdom, 1980.

[15] McWherter D., Peabody M., Shokoufandeh A., and Regli W. C.. Solid model databases: Techniques and empirical results. ASME/ACM Transactions, The Journal of Computer and Information Science in Engineering, Vol. 1, 2001. pp 300-310.

[16] Novotni M. and Klein R.. Shape retrieval using 3d zernike descriptors. Computer Aided Design, Vol. 36 2004, pp 1047-1062,

[17] Osada R., Funkhouser T., Chazelle B., and Dobkin D.. Shape distributions. ACM Transactions on Graphics, Vol. 2, 2002, pp 807-832.

[18] Shah J., Anderson D., Kim Y. S., and Joshi S.. A discourse on geometric feature recognition from cad models. ASME Transactions, the Journal of Computer and Information Science in Engineering, Vol. 1, 2001, pp41-51.

[19] Snead. C. S. Group Technology: Foundations for Competitive Manufacturing. Van Nostrand Reinhold, New York, 1989.

[20] Thompson W.B., Owen J., de St. Germain H., Stark Jr. S., and Henderson T.. Feature-based reverse engineering of mechanical parts. IEEE Transactions on Robotics and Automation, Vol. 12, 1999, pp 57-66.