CATEGORISATION OF SHAPES USING SHAPE FEATURES

SOO-HOON PARK AND JOHN S. GERO
Key Centre of Design Computing and Cognition
Department of Architectural and Design Science
University of Sydney

Abstract. We introduce a shape categorisation method for architectural drawings that takes shape characteristics derived from those drawings as features. We propose a feature-based shape analysis procedure and a formal model of shape categorisation based upon the measurement of featural similarity of shapes. Shape categorisation is demonstrated in experiments on architectural drawings.

1. Introduction

One of the difficulties for current computer-aid design packages is the extraction of design significances and semantics from drawings. This lack hinders the wider use of computers in the later processes of designing. This paper presents a shape classification and categorisation paradigm that has the possibility to be developed into a computational design aid as a semantic extractor and design characterisation tool. This paper presents an approach to the sorting and selection of figures in architectural drawings so that provided with the semantic description of design requirements, a system could suggest a group of drawing candidates from an image database with the shape characteristics most likely to satisfy the semantic design requirements.

One of the primitives comprising architectural drawings is “shape”. A shape is one of the most abstract design ingredients, to which designers assign various aspects of design information during a design process. Designers perceive, think, and manipulate shapes to suit their purposes. The essential element of a shape in 2-D or 3-D is a line. A shape normally refers to any finite arrangement of lines (straight, curved, open or closed) in the plane drawn in a finite area in a finite amount of time, which has a pictorial specification (Stiny, 1975; 1980). However, we distinguish a curve, which is an open arrangement of rectilinear or curvilinear lines, from a shape, which is a finite composition of closed and connected lines (Gero and Park, 1997; Park and Gero, 1999). This paper presents a computational method that develops a link between physical shapes and abstract concepts called features (Jared, 1984; Stiny, 1989; Shah, 1991). Features have been extensively used in geometric design assessment, especially in mechanical engineering, where feature recognition from geometric CAD representations has been a critical topic for implementing feature-based CAD systems (Meeran and Pratt, 1993; Brown et al., 1995; Tombre, 1995). This paper presents a rigorous feature-based shape analysis procedure for architectural drawings.

2. Feature-based Shape Analysis Procedure

This paper proposes a shape analysis procedure resulting in shape categorisation. This feature-based shape analysis procedure systematically computes all the necessary ingredients for shape categorisation, which include the encoding of shapes, shape feature detection, matching the structural shape description to shape semantics by feature classification, similarity measure for individuals, category feature lists, shape category definition, and shape categorisation. Shape feature classification and shape categorisation provide sufficient data for shape comparisons.

2.1 FEATURE-BASED SHAPE ANALYSIS PROCEDURE

Our feature-based shape analysis procedure consists of four discrete sequential processes. These processes are: *shape encoder, feature detector, feature classifier* and *shape analyser* as shown in Figure 1.

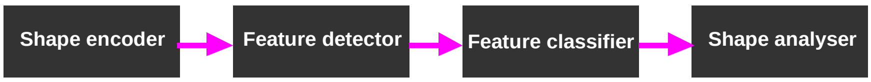


Figure 1. Shape analysis procedure

Each process has specific representation schemes, search algorithms, and necessary definitions and knowledge, in order to process data and to produce the required output. Data for each process are shapes, qualitative shape representations, called Q-codes[1], identified shape features and similarity measure to categories. The shape encoder takes shape images

[1] Q-codes have been developed as a qualitative shape encoding scheme, which converts shapes into systematically constructed qualitative symbols. Four separate Q-codes – A-, L-, K- and C-code – encapsulates shape characteristics in terms of angle, relative length, curvature and convexity (Park, 1999).

and transforms them into syntactic shape representations using the Q-code encoding scheme (Gero and Park, 1997). The feature detector takes this syntactic data and searches all the shape features that are identified in the form of structures in the Q-code encodings, called Q-words. The feature classifier takes the shape features as data, classifies them into groups and categorises shapes using definitions of categories. It further measures how typical the shapes are to the generic categories using a similarity measure. The shape analyser assesses shapes for how strong their characteristics are to the categories and compares shapes. Each process is explained below.

2.1.1 The shape encoder

The shape encoder converts a visual image into symbolic data. This process includes image vectorisation and Q-code conversion. It converts raster graphics to vector graphics and then converts shape contours into a set of Q-code encodings. The vector graphics contain information about vertices of the shape contour.

2.1.2 The feature detector

The feature detector takes the Q-code encoding as data and extracts all the shape features. This is a shape feature extraction process. After a shape is represented as a string of Q-codes, that is a Q-sentence[2], the feature detector looks for syntactic shape features as Q-words from the Q-sentence. The feature detector produces a group of identified shape features that are sorted according to Q-code lengths.

2.1.3 The feature classifier

The feature classifier takes shape features as data, classifies them into specified shape feature classes and measures the categorical typicality of each shape to the shape categories. The feature classification process results in several groups of shape features sharing commonalities and regularities based on the definition of shape feature classes. We choose basic shape feature categories according to shape characteristic classes in terms of repetition and convexity, which best distinguish shape feature characteristics in syntactic patterns. They are “iteration”, “alternation”, “symmetry”, “indentation” and “protrusion” categories. Iteration refers to a repetition of patterns with no interval; alternation refers to a repetition of patterns with regular or irregular intervals; and symmetry refers to a reflective arrangement of patterns.

---

[2] Q-sentence is one of the conceptual units of Q-code chunking. Linguistic terms are used for these units such as Q-code, Q-word, Q-phrase, Q-sentence and Q-paragraph. Q-sentence refers to a closed and complete contour of a shape (Gero and Park, 1997).

2.1.4 The shape analyser

The shape analyser compares shapes based on categorical knowledge of shape features. It examines a shape or a group of shapes to see how strongly a shape is associated to a category. Based on definitions of categories and classified shape features, the shape analyser examines shapes and determines their categorical prototype and average exemplar, and computes categorical membership from a group of shapes.

Shape categorisation is a result of shape analysis as well as the start of another analysis problem because it produces an explanation for a particular categorisation result. The explanation for shape categorisation provides the information for common or regular categorisation patterns that can be stored as new categorical shape knowledge.

2.2 CLASSIFICATION OF SHAPE FEATURES

Considering the classification task as an assignment of data to a predefined set of solutions on the basis of the object data (Stefik, 1995), the matching, or assignment, of syntactic descriptions to semantics involves several fundamental issues. The classification rules for matching is based on probabilities. When the rule selects a single optimal solution from two or more conflicting feature classes, the feature classifier should be able to select the appropriate solution with a higher probability. The second issue is the abstraction for appropriate class matching. The abstraction is a reduction of a description to a more general case. We try to generalise the syntactic data into a generic and predefined definition of shape feature class that corresponds to a particular semantic solution that also consists of abstract semantic descriptions and individual refinements such as shown in Figure 2.

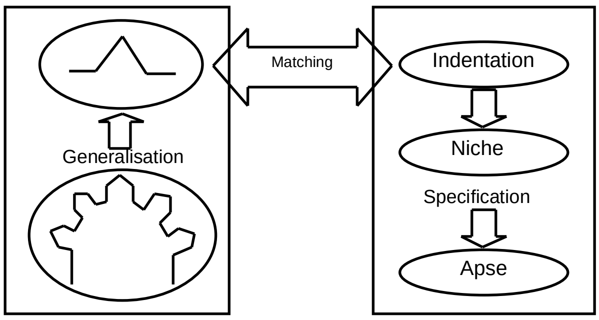


Figure 2. Architectural example of shape feature classification

The third issue in classification is that it uses a search model to examine the data and solution spaces. The feature classifier, given the data and the structured solutions, is provided with a method that matches data to solutions capable of handling several problems such as ambiguous data, missing data, and the reliability of data.

2.3. CATEGORISATION BASED ON SHAPE FEATURE CLASSES

The categories form the theoretical basis for shape comparison such that the shape semantics determine the assessment of the design ideas in shapes. The shape feature-based categorisation and the shape analysis is performed at the semantic level using a semantic feature analysis method (Pittelman et al., 1991; Estes, 1994). The semantic feature analysis method has been widely used particularly in the analysis of new vocabulary learning (Pittelman et al., 1991). We apply this method to the analysis of shapes. When a cognitive system encounters new concepts, it normally uses a certain cognitive structure called categories (Rumelhart 1980). The category structure associates verbal and visual concepts and serves as a semantic framework for related concepts stored in memory. Consequently the information about a newly encountered object is categorised and re-categorised and is integrated with the already-stored concepts (prior knowledge) so that it classifies the new object correctly. Then, the system discovers a way to differentiate the types of a new object − if it is alike or different from a certain object class.

The system relates new knowledge to the existing knowledge by classifying the semantic features in order to determine the similarity of objects. We find this method efficient in demonstrating the relationships among different concepts of unique objects. We need sufficient semantic precision among similar concepts because comparing different shapes will eventually show different semantic feature patterns when enough features are considered. Issues in semantic shape feature analysis include: shape categories, number of shape features, similarity measure, and inference on correct shape semantics.

3. Feature-Based Categorisation Model

3.1. CATEGORY AND SIMILARITY MEASURE

The “category” or “concept” is considered the essential cognitive element of our understanding of the world because we understand objects not at the individual level but at the class level in terms of categories or concepts. Categories always exist when two or more objects are treated equivalently. A category is thus an abstract organisational concept that groups individual objects described in terms of attributes.

A feature-based category model examines exemplars by the occurrences of their features; and a categorical comparison of exemplars is based on measuring the similarity in terms of features. Similarity between two objects is measurable in terms of proportions of common elements taken as shape features here (Estes, 1994). The similarity measure counts matching and mismatching semantic shape features and computes a numeric measurement based on this. The basic similarity measure equation as suggested by Estes (Estes, 1994) is:

$$Similarity\ (A,B) = s^{N-k} \tag{1}$$

Equation (1) computes an object's similarity according to semantic concepts by counting only mismatching shape features and by assessing how distantly shapes are related semantically. This equation is modified in order to infer a precise similarity measure by including mismatches (Park and Gero, 1999).

$$Similarity\ (A,B) = t^{r} \times s^{N-k} \tag{2}$$

There are five concepts involved in computing similarity. They are, *k:* the number of matching feature types; *N:* the total number of shape feature types; *r:* the occurrences of the feature; *s:* the value for mismatching features where $0 < s < 1$; and *t:* the value for matching features where $t \geq 1$.

3.2. FEATURE-BASED SIMILARITY MEASURE

The first task for the formal model of a shape exemplar-categorisation model is to specify the method of computing similarity measure for shapes, shape feature classes and shape categories. We use both the binary-valued feature identification and the production rule methods (Estes, 1994). Measuring similarity starts with the counting of shape feature occurrences to assess the commonality for the members. The simplest similarity for objects sharing a single common shape feature is measured by the following equation (single feature similarity):

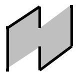
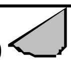
$$Similarity = t^{r} \times s \tag{3}$$

In equation (3), the shape features are counted in terms of matching "*t*" and mismatching "*s*" of a feature according to its occurrence "r".

Table 1 shows a simple illustration of similarity measure for the comparison of four arbitrary individuals {S1, S2, S3} with four arbitrary features {F1, F2, F3}. Individual characteristics in terms of matching (t ≥

1) and mismatching ($0 < s < 1$) features are reflected in similarity measure.

TABLE 1. Semantic feature analysis grid for simple shapes

Shapes / Shape features	 S1	 S2	 S3
F1 (sharp teeth) 	3	4	6
F2 (reflective symmetry)	0	0	1
F3 (rotational symmetry)	0	1	0

The similarity measure between two distinct shapes can also be assessed at the shape category level. The category level similarity is measured by summing all the similarity measures between one exemplar and all the other members of the category. The similarity measure at the category level for each exemplar is thus given in equation (4) (similarity to category):

$$\textit{Similarity-to-Category}\ (A) = \Sigma\ \textit{Similarity}\ (A,X) \tag{4}$$

"X" indicates all the members (exemplars) including the exemplar *"A"* in a shape category. The similarity measure for two exemplars at the category level results in a table where each exemplar is compared one by one and the individual comparisons are added to compute the categorical similarity. The result of measuring the similarity to category for those shapes in Table 2 is shown in Table 3. It measures by assigning "t" and "s" values to the equations (*see* Section 3.3 for details).

TABLE 2. Similarity measure of arbitrary individual shapes

	F1	**F2**	**F3**	**Similarity measure**
S1	t^3	s	s	t^3s^2
S2	t^4	s	t	t^5s
S3	t^6	t	s	t^7s
Sim (S1,S2)	t^3	t	s	t^4s

TABLE 3. Similarity measures of arbitrary individuals to category (t = 1.2, s = 0.8)

	S1	**S2**	**S3**	**Similarity to category (Total = 19.1)**
S1	t^5	t^4s	t^4s	$t^5 + t^4s + t^4s$ (5.8)
S2	t^4s	t^6	t^4s^2	$t^4s + t^6 + t^4s^2$ (6.0)
S3	t^4s	t^4s^2	t^8	$t^4s + t^4s^2 + t^8$ (7.3)

Assuming that a shape category "Cat-1" is represented by the exemplars "S1", "S2" and "S3" in Table 3, we can assess how strong each exemplar is associated to the category. One of the methods to determine the associative strength between an exemplar and a category is the measure of relative typicality (Estes, 1994). The relative typicality to category (Relative-typicality) is measured for each exemplar as a ratio:

$$Relative\text{-}typicality(Ex) = \frac{Similarity\text{-}to\text{-}Category(Ex)}{\Sigma Similarity\text{-}to\text{-}Category(X)} \tag{5}$$

"X" indicates all the members of the category including the exemplar *"Ex"*. The ratio measure for relative typicality for the previous exemplars {S1, S2, S3} of a category, for instance, is shown in Table 4.

TABLE 4. Relative typicality to a category

Exemplars	Relative typicality
S1	*Relative typicality* (S_1) = 5.8 / 19.1 = 0.30 (30%)
S2	*Relative typicality* (S_2) = 6.0 / 19.1 = 0.26 (31%)
S3	*Relative typicality* (S_3) = 7.3 / 19.1 = 0.20 (38%)

The relative typicality in Table 4 shows the associative strength of each exemplar to a category. For instance, we could say that the exemplar "S3 (38%)" is more strongly associated to the category than "S1 (30%)" so that we can say S3 is more typical than S1 to the category. This measure is used for the rating of the categorical typicalities of exemplars.

3.3 CATEGORY REPRESENTATION

A category can be represented in three ways: rule-based, prototype-based, or exemplar-based. The rule-based method defines a category by a set of necessary and sufficient conditions such that the categorisation task involves a process of testing whether an exemplar satisfies these conditions (Smith and Medin, 1981). Prototype theory considers the category to be represented by the prototype – the most dominant and typical member – and the categorisation becomes a process of comparing the exemplars to the prototype of a category (Rosch, 1973; 1978). The exemplar theory holds that a category is defined by the whole exemplars in the category and the categorisation is based on the similarity of a new object to all of the stored exemplars (Estes, 1986; 1994).

Firstly, a category can be represented by the deterministic selection rules with a set of necessary and sufficient featural conditions. The difference between necessary and sufficient conditions is the certainty factor of rules set with minimum and maximum certainties. The major criticism of the rule-based category representation (Ashby and Maddox 1998) is that the rules do not possess a graded structure distinguishing some members as more typical to the category than others. This is obviously not appropriate for the rating of categorical typicality. Secondly, a category can be represented by a prototype, which is the most typical and ideal member of the category. The assessment of category membership for a group of exemplars, therefore, becomes a matter of measuring their similarity to this prototype. It results in a gradation of category memberships. In our shape feature approach, we compute the similarity measure between an entity *"n"* and the prototype *"p"* by comparing the matching and mismatching features between an entity and the prototype. This similarity measure to the prototype is equivalent to categorical typicality measure in this case because the category is treated as equivalent to its prototype.

$$Typicality\ (n, p) = Similarity\ (n, p) = t^{r} s^{N-k} \quad (6)$$

A prototype representation of the category considers the ideal member for a category, such as the cathedral "Notre Dame", for instance, can be regarded as the representation for the category "Cathedral". The main criticism to this is that it loses all the information about the category and the various correlations of other members ignoring the contextual knowledge because only the prototype is considered as the representation of the category. Other members perform a critical role in category representation. This method is, however, useful in the category representation when we have a clear idea about the distinctive and well-known member for the category. Thirdly, the exemplar representation of a category considers all category exemplars that have been encountered and stored. We can determine the categorical membership of an entity by the similarity comparison of it to the representation of every exemplar of the relevant category (Estes, 1986; 1994). This category representation, thus, considers a collection of features of every exemplar. The categorical typicality for an entity is computed as the sum of the one-to-one similarity measure between the entity "*n*" and all the exemplars "*x*" as follows:

$$Typicality\ (n) = \Sigma\ Similarity\ (n, x) \quad (7)$$

The main advantage of the exemplar representation of category over the prototype representation is the sensitivity of a category definition covering the maximum available features. The main criticism also comes from the way a category is represented with every member (Ashby and Maddox, 1998). Firstly, it is problematic whether categorisation depends exclusively on exemplars rather than dominant features. Secondly, the category computation involving every exemplar of a relevant category is intuitively unrealistic especially for a category with very large members.

Consequently, the categorical membership for a new shape can be assessed in three possible ways: by checking the rules, by comparing the featural similarity with the prototype, or by examining the similarity with all the members in the category. For computational purposes, it is efficient to consider two types of shape feature lists: the feature lists of a new shape and the common feature list of the shape group for the categorical typicality measure. A category feature list could come from one member (prototype), from a small number of members (multiple prototypes), or from the whole of the members. Following this discussion, the categorical typicality measure equation can be redefined as follows (categorical typicality measure based on the common feature list):

$$Typicality\ (s) = Similarity\ (f_s\ , F) \tag{8}$$

Categorical typicality for a shape "*s*" to a certain category represented with the common feature list "*F*" is measured by the similarity between the individual feature list "f_s" from the shape "*s*" and the common feature list "*F*" from the category representation. Categorical typicality measures often come out as large numbers with large differences. This is not satisfactory for an overview of all members and it often misleads because of the large numeric difference. One of the ways to include an overview of the typicality measures is to compare them to that of the prototype so that we can have relative measures of categorical typicality ranging from zero to 100% to that of the prototype. We can set two basic constraints for the relative categorical typicality measure:

- t + s = *C* (where *C*, a constant, is a small number such as 2); and
- determine optimal t and s values such that the average of the categorical typicality measures to that of the prototype approaches 50%.

3.4 CATEGORISATION OF SHAPE EXEMPLARS

Our goal in feature-based shape categorisation includes the acquisition of the generic shape categories, the categorisation of shapes to existing categories, the determination of a new category, and the identification of key shape feature classes. We approach this in three ways to construct the new categories: pair-wise categorisation, key-feature-based categorisation and sub-categorisation.

- *Pair-wise categorisation:* We assume a shape group as members of a category from which we derive a category pair with contrasting characteristics so as to assign shape exemplars into these two categories.

- *Key-feature-based categorisation:* We can create a new shape category based on the key shape feature classes detected from the shape group.
- *Sub-categorisation:* We can create new categories by refining the existing category definitions combining two or more existing categories. This produces a series of subordinate categories.

4. Shape Categorisation

4.1 CATEGORISATION AS A RESULT OF SHAPE ANALYSIS

The primary instrument for the shape comparison has to be the similarity or the difference from which we assess the individual or the collection of shapes. The primary model of shape analysis and comparison has to explain explicitly the deterministic link between shape characteristics and shapes. This paper argues that a shape categorisation method is best suited for this purpose. Feature-based shape categorisation supports sufficient data for the comparison of shapes, as follows.

- *Feature-category connection*: It offers an explicit link between the featural description of shape characteristics and a shape group under a particular category.
- *Grouping of individual shapes*: It provides the principles of grouping individual shapes based upon shape characteristics.
- *Sub-group*: It can split a shape group into sub-groups according to associative strength to many categorical concepts.
- *Assessing similarity*: It can assess the shape similarity and cope with shape groups sharing equivalent or common shape characteristics.
- *Novel view*: It provides multiple views to a shape group with new ways of dividing the shape group.
- *Exhaustive search for groupings*: It is capable of searching all the possible groupings of shapes based on shape characteristics.
- *Ordering of groups*: It is capable of sorting out the refined groups by their categorical typicality.
- *Explanation for grouping*: It provides explanations for the grouping results with relevant shape features and semantic labels.

Figures 3 (a), (b) and (c) show the major tasks of this shape categorisation.

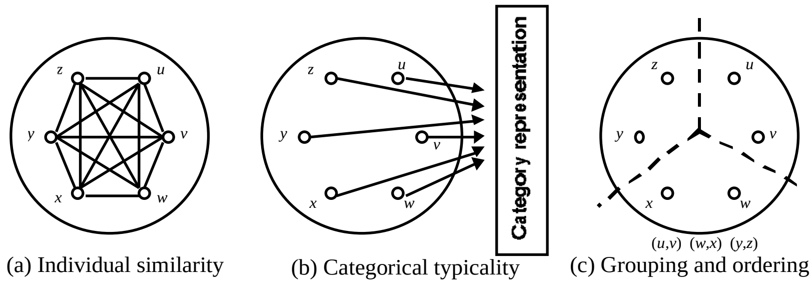


Figure 3. Three patterns of shape categorisation

4.2 SHAPE CATEGORY EXAMPLES

For the categorisation experiment, we propose simple shape categories composed of a square ***(SQR)*** and three protrusions *(P)*. These are bounded by the finite shape exemplars sharing featural commonalities, whose members are the most representative and regular shapes for the categorical characteristics.

SQR+3P category This shape category includes the shape exemplars with a square (*SQR*) and three protrusions (*3P*) on the edges as members. There are five members in this category as shown in Figure 4.

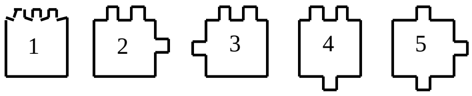


Figure 4. Shape category SQR+3P

SQR+2P+1C This shape category includes the shape exemplars with a square (*SQR*), two edge protrusions (*2P*) and one corner protrusion (*1C*) features. The membership is given to the following exemplars in Figure 5.

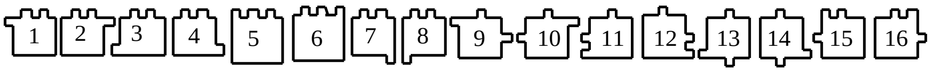


Figure 5. Shape category SQR+2P+1C

SQR+1P+2C The shape features characterises this category include a square (*SQR*), one edge protrusion (*1P*) and two corner protrusions (*2C*). The exemplars are shown in Figure 6.

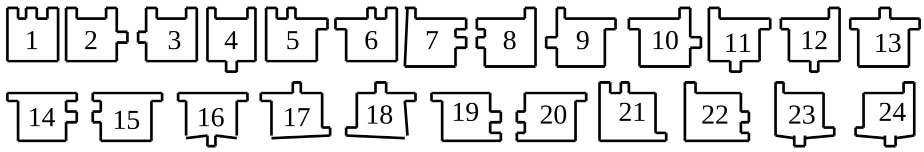


Figure 6. Shape category SQR+1P+2C

SQR+3C This shape category is characterised by the shape features of a square (*SQR*) and three corner protrusions (*3C*). The membership includes the following exemplars in Figure 7.

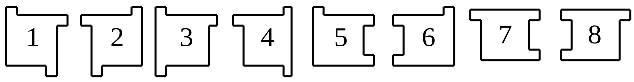


Figure 7. Shape category SQR+3C

4.3 EXPERIMENT 1: CATEGORY DETERMINATION FOR A TEST SHAPE

The first experiment considers a single test shape and determines if it belongs to an existing shape category considering a single shape and a single category. Shape categories are represented in three ways: based on rules, prototype, or exemplars. Experiment 1 takes the test shape in Figure 8 and compares it to the category *SQR+3P*.

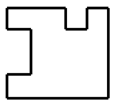


Figure 8. A test shape

The rule-based representation of category *SQR+3P* states the test shape is not a member of the *SQR+3P* category because it does not match all the necessary and sufficient conditions. It provides a straightforward answer to the membership question but it produces no additional information. The prototype representation of the category *SQR+3P*, based on S4, shows the category membership test result in Table 5.

TABLE 5. Categorical similarity using the prototype-based method (t = 1.01, s = 0.99).

Similarity(test shape, prototype)	$t^{34}s^{232}$ (0.14) → **27**%
Similarity(S1, prototype)	$t^{84}s^{189}$ (0.35) → **35**%

The test shape shows 27% similarity that is lower than the least similar shape S1 (35%) to the prototype S4. The test shape, therefore, is not considered to be the member of the category *SQR+3P* although showing a close similarity. Using the exemplar-based method the similarity measures of the test shape are 90.6 (commonality) and 0.46 (difference). The similarity measures of the members (S1, S2, S3, S4, S5) range from 1502.6 to 2409.3 (commonality) and from 1.12 to 1.18 (difference). The test shape shows a lower similarity in commonality and less difference to the category. Higher similarity measure in commonality between test shape and a member of the category can be interpreted in two contrasting ways: (1) both contain many common features, or (2) both contain many unique features against the categorical feature list. Since this test shape shows very little commonality, exemplar-based category excludes this shape as a member of the category *SQR+3P*.

4.4 EXPERIMENT 2: CATEGORISATION OF MULTIPLE NEW SHAPES

The next experiment concerns the categorisation of a group of new shapes that are not categorised in the existing shape categories.

4.4.1 Categorical typicality of new shapes to the existing categories

This experiment examines, firstly, if any test shape belongs to one of the four existing categories and, secondly, if the test shape group forms a set of discrete groupings as separate shape categories. The test shapes in Figure 9 are constructed as the counterpart shapes for the members in the four categories in Figures 4, 5, 6 and 7. We deliberately convert a shape with a square and three edge protrusions into the counterpart shape as a square and three edge indentations. These shapes are actually members of categories *SQR+3I* (S1–S5), *SQR+2I+1CI* (S6–S13), *SQR+1I+2CI* (S14–S19), and *SQR+3CI* (S20) in Figure 10. This experiment tests to see if category theory identifies the members and new or unique categories different from the existing categories. The first step is to examine each shape in Figure 9 to determine if any of the shapes is categorised into one of the four existing categories.

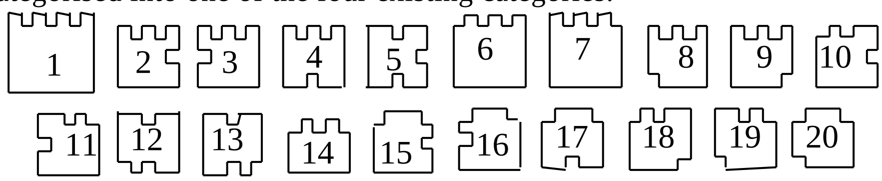


Figure 9. Test shape group

The categorical typicalities of indentation-based shapes are measured to the four existing protrusion-based categories. The membership is given to shapes when the test shape's typicality measure is larger than the lowest categorical typicality of any existing protrusion members in Figures 4, 5, 6, and 7. The result shows that only test shape S17 in Figure 9 is considered a member of category *SQR+1P+2C*, meaning that a particular square and one-edge and two-corner indentation shape description overlaps with that of a square and one-edge and two-corner protrusion shape description.

The second step is the categorisation of the test shapes to new categories, where we expect four new indentation-based categories be produced as a result. We approach the definition of new categories for a group of test shapes by classifying them into the five generic shape feature categories, namely indentation, protrusion, iteration, alternation, and symmetry categories, which provide filters that sort our particular shape features from the whole feature list.

4.4.2 Categorisation based on indentation

Two types of indentation features are discovered and each shape shows three occurrences of indentation features. The result matches our expectation regarding the four indentation-based categories. The categorisation result is shown in Figure 10.

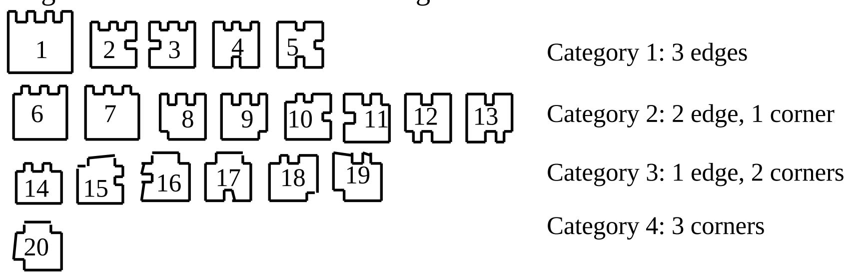


Figure 10. Indentation-based categorisation of test shapes

The first goal was to check membership of an existing shape category. One of the test shapes, S17 in Figure 9 was shown to belong to one of the existing protrusion category, SQR+1P+2C. The second goal was to construct new categories out of the test shape group. For this task, we used four generic shape categories, in terms of indentation, protrusion, iteration and symmetry, to filter out the category feature lists sharing commonality. Consequently, we have four filtered lists of shape features that form an exemplar-based category representation of indentation, protrusion, iteration and symmetry categories. We achieved two types of categorisation results. One is a list of feature–categorisation links and the other is a categorical typicality measure for those generic shape categories. The feature–categorisation result tells us the following.

· *Possible shape groupings*: This is the result of an exhaustive search for possible shape groupings. Any attempt to regroup test shapes under the generic shape categories falls into the combination of sub-groups of this categorisation result.
· *Ordering of features and groupings*: Feature–categorisation results can be ordered in several ways according to the number of shape features, the length of features, the number of groupings, and the length of groupings. Each ordering unveils an important structure

from which we could access an image database of the test shapes such that we could select shapes using abstract and qualitative predicates distinguishing particular characteristics of the ordered structure.

- *Explanation*: It provides the explanation for particular shape groupings. Given a particular shape grouping for a test group, the reason for the decision is supported by feature–categorisation connection lists.

The validity of the categorisation based on generic feature categories is confirmed by the results of the categorical typicality measure, which assesses the individual shapes in terms of generic shape categories. The measure provides us with an abstract insight over the individual as well as categorical characteristics. This becomes explicit when we compare the categorical typicality measure of each test shape using qualitative symbols.

5. Categorisation of Architectural Drawings

The whole shape group is considered as members under the category labelled anonymous. Individuals, in this case, are initially compared to the group characteristics so that characteristic shapes can be selected for that shape category. A shape category, under this shape analysis method, is defined by the 'common feature list' that is constructed from the shape features of all the members in this anonymous category. The common feature list is composed of shape features from at least two members in the group. This method of common feature selection greatly reduces the number of features for the anonymous category definition.

5.1. CATEGORISATION EXPERIMENT USING FIVE GENERIC CATEGORIES

Figure 11 shows a group of Alvar Aalto's plan drawings on which we will perform a categorisation experiment. The shapes are presented in the order of contour complexity, which is determined from the addition of Q-code length and the number of protrusions. Twelve shapes in Figure 11 are encoded in A- and L-codes for the contours of the shaded boundaries (Park and Gero, 1997). The generic shape categories are filtered out to form specific category feature lists, on which each shape is assessed in terms of categorical typicality.

Firstly the whole shape group is considered as members under the category label of arbitrary shape characteristics. Individual shapes, in this case, are initially compared to the group characteristics so that distinctive shapes can be selected for that shape category. A shape category, under this shape analysis method, is defined by the 'common feature list' that is constructed from the shape features of all the members in this anonymous category.

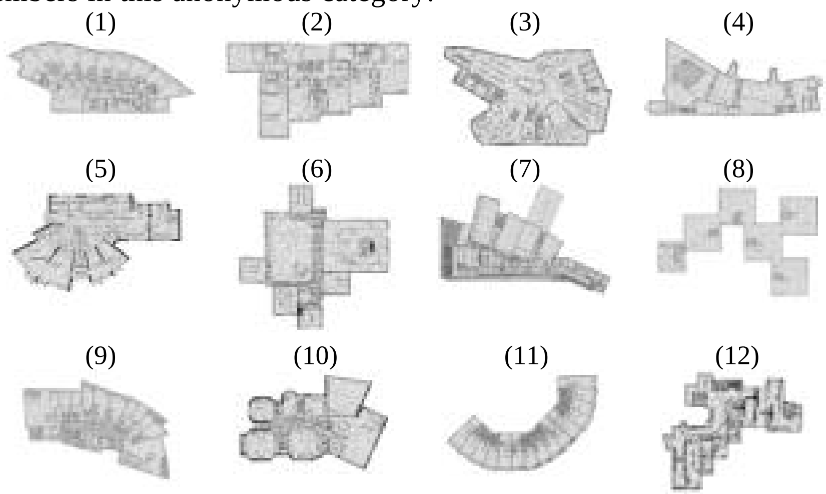
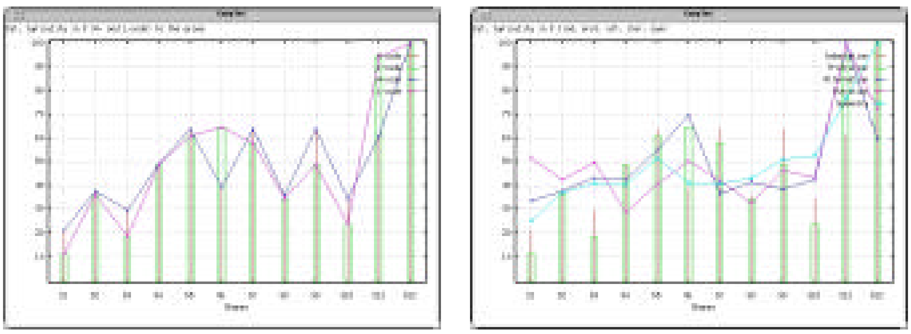


Figure 11. Architectural plan drawings by Alvar Aalto (Schildt, 1994)

Figure 12(a) shows the result of preliminary measurement of categorical typicality for the twelve shapes. The categorical typicality measure clearly shows which shape contains most of the common shape features and feature diversity. Those shapes containing most of the common shape features are considered to be the prototype that it is the most representative of the shapes regarding shape characteristics as a group.



(a) (b)

Figure 12. Categorical typicality measure

Pair-wise categorisation separates the categorisation results into two contrasting groups. This is a primary categorisation because this method does not depend on any prior knowledge or representations of specific shape categorisations or shape features. This would be the first step in examining a group of shapes since it provides a way to pre-inspect the overall shape characteristics without further details. Pair-wise categorisation produces the results as "{6, 10, 11, 12}:{1, 2, 3, 4, 5, 7, 8, 9}" and "{11, 12}:{1, 2, 3, 4, 5, 6, 7, 8, 9, 10}" for A-code and L-code encodings respectively, where the distinction is made at the 50% point to that of the most typical member. The labelling of these pairs of shape categories can only be generalised as "typical" or "non-typical" distinctions. The result shows that the common members for each group are "{11, 12}:{1, 2, 3, 4, 5, 7, 8, 9}". Higher occurrences of these shape features indicate more distributed occurrences of general features and less occurrences of idiosyncratic features. This also means that those members in the non-typical categories are considered to be simpler and regular in their composition of shape contours.

5.2 KEY-FEATURE BASED CATEGORISATION

Indentation- and protrusion-based categorisation: the shape analysis system detects three types of indentation features and five types of protrusion features in the primary granularity of A-code encodings. Vertical lines in Figure 12(b) show the categorical typicality measure to indentation and protrusion categories. The pair-wise categorisation results are "{5, 7, 9, 11, 12}:{1, 2, 3, 4, 6, 8, 10}" for indentation and "{5, 6, 7, 11, 12}:{1, 2, 3, 4, 8, 9, 10}" for protrusion.

Alternation-, iteration- and symmetry-based categorisation: the second group of generic categories includes alternation, iteration and symmetry categories that are represented with relevant shape features, for which the shape analysis program detects 1281, 128 and 79 shape features of A- and L-codes for each shape category. Line graphs in Figure 12(b) show categorical typicality measure of the twelve shapes to each category. Pair-wise categorisation results show that shapes{5, 6, 11, 12} for alternation, shapes {1, 6, 11, 12} for iteration and shapes {5, 9, 10, 11, 12} for symmetry categories are identified as typical exemplars.

5.3 INTERPRETATION OF CATEGORISATION RESULTS

The experiment results in classifications of drawings under several discrete shape categories. Table 6 shows the pair-wise categorisation for each shape. The shapes are listed in the order of categorical typicality. Table 6 shows a simplified assessment of categorisation results. The result shows some of the shapes always contrast to each other for feature-based groupings: "{11, 12}:{2, 3, 4, 8}" shown as shaded patterns, and occur in all cases. Some of the shapes show a strong tendency to be a contrasting pair: “{5, 6}: {1, 7, 9, 10}”. Shapes {11, 12} are selected as the most noticeable and representative exemplars for every categories except A-code group category, which selects shape {6}.

TABLE 6. Strong and weak tendencies of drawings to shape categories

Cat	**Strong tendency**						**Weak tendency**									
Grp-A	6	11	12	10		:	5	8	9	4	7	3	2	1		
Grp-L	11	12				:	8	7	5	3	10	4	9	2	6	1
Ind	12	5	7	9	11	:	4	6	2	8	10	3	1			
Prot	12	11	6	5	7	:	9	4	2	8	10	3	1			
Alt	11	6	12	5		:	3	10	4	8	9	2	7	1		
Iter	11	12	1	6		:	3	9	10	2	7	5	8	4		
Sym	12	11	10	5	9	:	8	6	7	3	4	2	1			

6. Discussion

We have explored, in this paper, a feature-based shape analysis system and formal shape categorisation model including category representation methods and tools to measure similarity of shapes based upon shape characteristics. We have demonstrated how shape categorisation can be applied in analysing and grouping shapes in architectural drawings.

We are able to acquire relevant information for characterising architectural drawings in terms of individual or group characteristics. For the basic categorical shape comparison these results could be used to discover a stylistic consistency for drawings from a specific designer for a particular period of time. This analytic is thus suitable to be developed into a characterising tool for architectural drawings. It could produce analytic results as data for the following possible tasks: to check the consistency and uniformity of style from a set of drawings; to check the variations of style from a set of drawings; to examine the stylistic consistency from drawings across a group of designers; to explore the stylistic changes of a particular designer’s works across time; and to possibly predict future stylistic changes of a designer based upon the regularities on the style change found in the analysis data. Further, it has the potential to be used as a content-based indexing system for drawing databases.

This system has the capacity to be further developed into a sketch-based design aid tool that searches similar pictorial data using intuitive and undetailed sketch input. It may possibly be used as an abstract-shape synthesis system by reversing the shape analysis procedure to produce Q-code based preliminary shape descriptions as design alternatives for a particular design problem.

Acknowledgements

This research has been supported by an Overseas Postgraduate Research Award and by a grant from the Australian Research Council.

References

Ashby, F. G. and Maddox, W. T.: 1998, Stimulus categorization, *in* M. H. Birnbaum, (ed.), *Measurement, Judgment, and Decision Making*, Academic Press, San Diego, 251-301.

Brown, K. N., Sims Williams, J. H. and McMahon, C. A.: 1995, Conceptual geometric reasoning by the manipulation of models based on prototypes, *AIEDAM* **9**, 367-385.

Estes, W. K.: 1986, Memory storage and retrieval processes in category learning, *Journal of Experimental Psychology*, **115**, 155 – 174.

Estes, W. K., 1994, *Classification and Cognition*, Oxford, New York.

Gero, J. and Park, S.-H.: 1997, Computable feature-based qualitative modeling of shape, *in* R. Junge (ed.), *CAAD Futures '97*, Kluwer, Dordrecht, 821-830.

Jepson, A. D. and Richards, W.: 1993, What makes a good feature?, *in* L. Harris and M. Jenkin (eds), *Spatial Vision in Humans and Robots*, Cambridge, Cambridge, 89-125.

Meeran, S. and Pratt, M. J.: 1993, Automated feature recognition from 2D drawings, *CAD* **25**(1), 7-17.

Park, S.-H. and Gero, J. S.: 1999, Qualitative representation and reasoning about shapes, *in* J. S. Gero and B. Tversky (eds), *Visual and Spatial Reasoning in Design*, Key Centre of Design Computing and Cognition, University of Sydney, Sydney, Australia, 55-68.

Park, S.-H.: 1999, *Qualitative Representation and Feature-based Analysis of Shape*, PhD thesis, University of Sydney.

Pittelman, S. D., Heimlich, J. E., Berglund, R. L. and French, M. P.: 1991, *Semantic Feature Analysis*, International Reading Association, Newark.

Rosch, E.: 1973, Natural categories, *Cognitive Psychology*, **4**, 328-350.

Rosch, E.: 1978, Principles of categorization, *in* E. Rosch, and B. B. Lloyd, (eds), *Cognition and Categorization*, Erlbaum, Hillsdale, 27-48.

Rumelhart, D. E.: 1980, Schemata: The building blocks of cognition, *in* R. J. Spiro, B. C. Bruce, and W. F. Brewer (eds), *Theoretical Issues in Reading Comprehension*, Erlbaum, Hillsdale, 33-58.

Schildt, G.: 1994, *Alvar Aalto: The Complete Catalogue of Architecture*, Design and Art, Rizzoli, New York.

Shah, J. J.: 1991, Assessment of features technology, *CAD* **23**(5), 331-343.

Stefik, M.: 1995, *Introduction to Knowledge Systems*, Morgan Kaufmann, San Francisco.

Smith, E. E., and Medin, D. L.: 1981, *Categories and Concepts*, Harvard University Press, Cambridge.

Stiny, G.: 1975, *Pictorial and Formal Aspects of Shape and Shape Grammars and Aesthetic Systems*, Ph.D. Thesis, University of California, Los Angeles.

Stiny, G.: 1980, Introduction to shape and shape grammars, *Environment and Planning B*, **7**, 343-351.

Stiny, G.: 1989, Formal devices for design, *in* S. L. Newsome, W. R. Spillers and S. Finger (eds), *Design theory '88*, Springer-Verlag, New York, 173-188.

Tombre, K.: 1995, Graphics recognition – general context and challenges, *Pattern Recognition Letters*, **16**, 883-891.

This is a copy of the paper: Park, S-H. and Gero, J. S. (2000) Categorisation of shapes using shape features, *in* Gero, J. S. (ed.), *Artificial Intelligence in Design'00*, Kluwer, Dordrecht, pp.203-223.