# ANALYSIS OF ARCHITECTURAL SKETCHES USING CATEGORICAL SHAPE KNOWLEDGE BASED ON SHAPE FEATURES

SOO-HOON PARK AND JOHN S. GERO
*Key Centre of Design Computing*
*Department of Architectural and Design Science*
*University of Sydney, NSW, 2006, Australia*

**Abstract.** Shape feature analysis method is suggested as a computational support for the association of pictorial patterns of sketches with design semantics. Geometric patterns have been represented with qualitative scheme which is capable of representing classes for a collection of instances. Similarities to the particular shape feature categories have been measured to compare the sketch instances.

## 1. Design Ideas in the sketch

### 1.1 MODELING OF AMBIGUOUS 2-D SHAPE

Modeling and reasoning about shapes are central activities in architectural design. In the early stages of designing, sketches play important roles in exploring design concepts. Those shapes in the sketches involve a class level of description in which a single shape description represents a collection of possible shape instances sharing similar descriptive characteristics. Modeling shapes at the sketch level thus requires the ability to represent the ambiguity, uncertainty, and incompleteness properties of conceptual shapes through abstract class definitions.

### 1.2 HANDLING SHAPES IN SKETCHES

As an easy and basic tool for designers in the conceptual stage, sketching shares those characteristics of the early stages of the design process (Coyne et al. 1990; Green 1992) which include transformation of design requirements into design specifications; making high-level qualitative design decisions related to functional performance; choosing classes of design components; and qualitative comparisons and selections of design alternatives.

CAADRIA '98 : Proceedings of The Third Conference on Computer Aided Architectural Design Research in Asia.
eds. T. Sasada, S. Yamaguchi, M. Morozumi, A. Kaga, and R. Homma
April 22-24, 1998. Osaka University, Osaka, Japan. Pp. 405-414

Sketching can be contrasted in many ways to final drafting. Sketching aims at a sufficient and minimal qualitative description rather than a complete quantitative description; describing design classes covering a range of individuals; and providing an abstract level of geometric and topological composition and explanation rather than precise drafting.

Design ideas are shared and communicated through the shapes and forms in sketches where ideas are expressed as characteristic pictorial patterns implying various design qualities (Alexander, 1979). Understanding the implications of those pictorial patterns is an important aspect of intelligent CAD system. Sketch modeling requires a modeling scheme that satisfies abstraction and class handling characteristics which cannot be sufficiently supported by conventional numeric modeling scheme. One of the suitable approaches that encapsulates geometric patterns into meaningful conceptual chunks is the feature concept which has been extensively used in mechanical engineering (Shah, 1991). This paper presents a qualitative modeling approach for the feature representation of shapes. It supports multiple levels of detail and symbolic pattern matching for geometric patterns (Gero and Park, 1997).

## 2. Shape Feature Analysis of Sketches

Sketch level drawings contain many preliminary expressions of design ideas and design semantic qualities more than the description level of, say, location of a pixel. Any shape representation for sketch analysis should provide the capability to group together and explicitly name characteristic sets of data or fragments of a shape. Explicit association of particular structures of shape patterns to names or labels such as “zig-zag” or “indentation” provides the key to identifying one shape from another at a semantic level. Shape feature analysis of sketches examines the shapes and identifies the types and occurrences of shape features to understand and determine design ideas underlying the pictorial configuration of sketches.

### 2.1 FEATURE BASED SKETCH ANALYSIS

This paper presents a feature based sketch analysis as a process comprising four phases: shape feature representation, feature recognition, feature categories, and comparison based on feature categories.

#### *2.1.1 Feature representation of a sketch*

This phase involves shape representation and encoding. A qualitative representation is applied to the shape in order to cover a range of design instances with a single representation. A qualitative coding scheme is used as the base representation.

*2.1.2 Feature recognition from the sketch*

This phase involves chunking the encodings into manageable groups and identification of the types and occurrences of shape features. A symbol sequence of a sketch encoding is subdivided into chunks using linguistic analogies. Basic chunks of encodings are recognised as “words” and their occurrences are counted.

*2.1.3 Categories of features*

Various shape features recognised from the sketch are grouped into categories. This paper examines shape feature categories in terms of syntactic regularity.

*2.1.4. Comparisons of shapes based on feature categories*

This phase uses similarity measure equations to compare either one shape with another shape or a group of shapes with shape feature categories. Shapes with known semantics can then be used as a benchmark to select the most similar design alternative for that category.

2.2 FEATURE ANALYSIS SCHEME OF SHAPE

*2.2.1 Feature Representation*

A feature-based design approach allows explicit semantic representation of geometric data (Jared 1984). One of the aims of feature-based design is to overcome the inability of the conventional geometric based representation schemes in associating the geometry with design meaning. Various feature extraction methods have been suggested to associate design semantics with geometric or topologic entities in feature models (Shah 1991), and one of them is syntactic pattern recognition. This paper presents a four phase syntactic feature representation scheme as shown in Figure 1.

*2.2.2 Physicality Symbolic representation (Q-code Encoding)*

The significant geometric and/or topological characteristics of shapes are examined and transformed into a symbol sequence at this stage. Three shape attributes – vertex angle (A), relative length of edge (L), and curvature (K) – are represented as symbols in Q-codes (Gero and Park, 1997), which produces a qualitative formulation with symbols and sign values (+, 0, -) ($\alpha$={A,L,K}). A symbol sequence of Q-codes functions as the syntactic representation of a closed and connected shape. The structure of the Q-code encoding syntax is further subdivided into smaller chunks of symbols named using a linguistic analogy. Some of those definitions are as follows:

Q-code, $\alpha$: $\alpha=\{\beta_i \mid \beta \;\; \{A,K,L\}, i \;\; \{+,0,-\}\}$
Q-word, $\omega$: $\omega_i \;\; \alpha_j$, $i=1, j \;\; [1,m]$ or $\omega=\alpha_j \ldots \alpha_k$, $|k-j|=i, i=[2,m]\}$

Q-sentence, $\sigma$: $\sigma=\{\alpha_1\ \alpha_2 \ldots\ \alpha_m\}$ where $m=length(\sigma)$

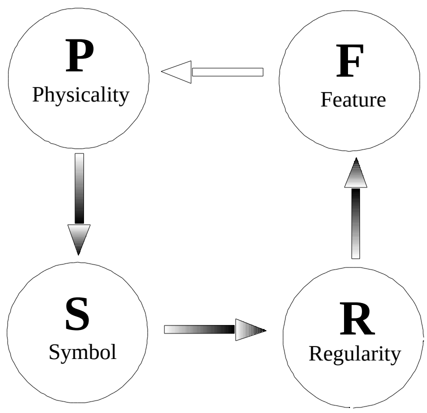


*Figure 1.* Four discrete phases of syntactic feature representation

### *2.2.3 Symbol Syntactic Reguarity*

Q-word is the smallest chunk of encodings that encapsulates the geometric and topologic characteristics of shapes in segments. A closed and connected shape produces a symbol sequence called a Q-sentence. The search for syntactic regularities from the shape representation is performed for the types and occurrences of Q-words in the Q-sentences using syntactic pattern matching.

When types and occurrences of Q-words in the shape are identified, the feature recognition system searches for syntactic regularities from the identified Q-words. In this paper, we look for two types of syntactic regularities – regularities in repeating syntactic patterns and regularities in syntactic patterns for known geometric pattern classes. Regularities in repetition can be one of alternation, iteration or (reflective) symmetry (Martinoli et al. 1988) categories. Regularities for known patterns, examined in this paper, are the most basic features – the indentation and protrusion categories.

### *2.2.4 Syntactic Regularity Feature*

Features encapsulate design significances in the segments of a geometric model and they associate those segment patterns with design semantics. The Q-code encoding scheme of feature representation describes physical characteristics in the symbolic representation scheme in such a way that a single symbol sequence in the qualitative formalism represents a group of geometric models sharing common shape attributes in a class. As a class representation, the syntactic regularity found in the symbol sequence can be associated with those geometric regularities with individual variations in numeric data. Some of geometric patterns match familiar verbal descriptions and others do not. Syntactic regularities are those geometric patterns that show significant structural information so those can be recognised as particular shape features with the ability to analyse the design models for their shape characteristics.

#### *2.2.5 Feature Physicality*

This mapping operates if the feature definition in a syntactic representation matches any geometric or topological patterns and checks if the symbol representation appropriately handles the class definition for a range of various individual geometric entities.

## 3. Sketch Analysis based on Categories of Shape Features

### 3.1 CATEGORICAL KNOWLEDGE OF SHAPE FEATURES

According to Rosch (Rosch et al. 1975; Rosch and Lloyd 1978), categories are represented logically by attributes in which membership is defined by a sample's possession of a simple set of critical features and instances possessing the same number of critical features have an equal degree of membership. Instances of the category are characterised into a prototype (clearest case, best example of the category) and nonprototype members and nonprototype members have tendency toward an order from better to poorer examples (Rosch et al. 1975).

This paper makes abstractions and comparisons of sketches primarily consisting of non-semantic descriptions on the basis of the definitions and the identifications of categories, categorical prototypes, and gradients of categorical memberships in terms of the frequency of shape feature occurrences. The associative strength between the geometry and the design semantics determines the gradients of category membership.

Classification of instances can either be based on the probability of valid shape features in the category (probability model) or be based on the matching to the prototype of the category (distance model) (Rosch 1975). Since categorical prototypes cannot precede category definition, this paper examines gradients of category membership before comparison with the prototype of the category. Shape feature categories are associated with labels (or design semantics) and they show inclusive relations between categories and their superordinary category. Five basic shape feature categories used are: "indentation category", "protrusion category", "alternation category", "iteration category", and "symmetry category". Among these, alternation, iteration and symmetry categories are shape feature categories in which repeating syntactic patterns occur, and they have a superordinate category: "repetition category".

### 3.2 SIMILARITY AMONG SKETCH ALTERNATIVES

Categorisation of shape features is more than just classification. Categorisation implies that knowledge of the category to which an object belongs tells us something about its properties (Estes, 1994). Shape feature categories are the results of generalisation and abstraction of a collection of objects and the basis of the generalisation is the similarity among those objects. According to Estes, similarity between any two objects or situations is measurable in terms of proportions of common elements (Estes 1994), and he suggested a similarity measure equation between two objects A and B as "$Similarity(A,B)=s^{N-k}$". We suggest a similarity measure equation as "$Similarity(A,B)=t^{r}s^{N-k}$" because Estes' equation do not count multiple occurrences of features where "t" = match (t>1), "s" = mismatch (0<s<1), "*r*" = number of occurrences of the feature, "*N*" = total number of shape feature types, and "*k*" = number of matching feature types.

As a basis for comparisons between sketches, similarity is measured in two ways: similarity to the category and similarity to the categorical prototype. We define categorical prototype as the example which has the most commonality to the category and the average examplar as the example with the average typicality to the category. Shape feature analysis of sketches adopts the exemplar-similarity model (Estes 1994) and looks for higher typicality in terms of commonality in shape features.

### 3.3 SIMILARITY MEASURE AND ANALYSIS OF SKETCHES

Figure 2 shows a sketch by Charles Correa (Lacy 1991). The sketch contains shape examplars (A ~ J). The analysis carried out is to select categorical prototype and average example for a given category definition and for a given prototype. Figure 3 shows the diagram that illustrates this process.

Samples from A to J are separate from each other and show various design solutions with different shape patterns. First, all the samples are examined according to the shape feature categories with explicit definitions. Measuring similarity to the category provide us data to select the category prototype which has the highest commonality of shape patterns in that feature category. This produces the category structure. Those discrete shape feature categories under consideration have been identified in terms of syntactic regularity which means that those sketches with the most orderly yet complicated patterns of structure show greater tendency to become the typical prototypes for those categories. The average examplars, thus, can be predicted to be those in the middle in terms of complexity and order. Second, one sample can perform as an arbitrary category prototype and other samples are compared to determine which show the most commonality to that sample. In this case the general syntactic regularity does not affect the selection of the sample, but the salient shape patterns in the typical examplar (which is the category prototype) perform as the selection factors.

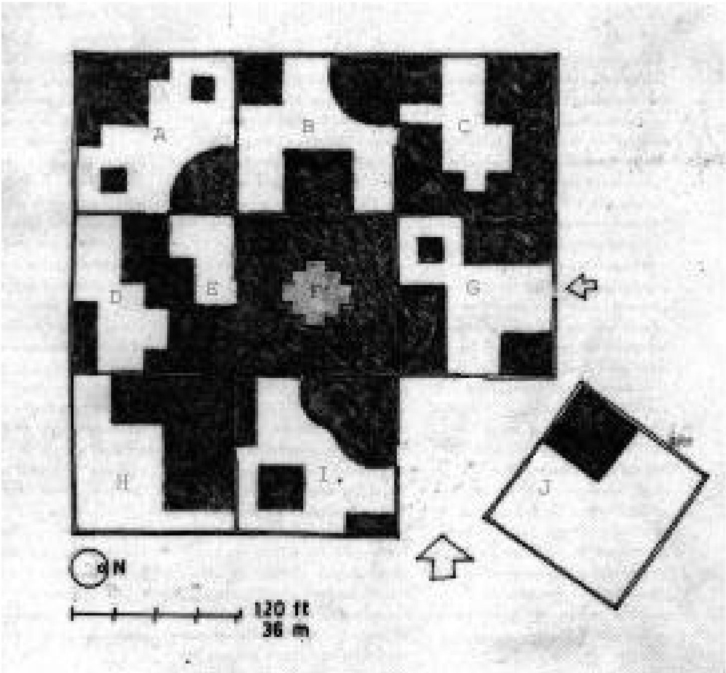


*Figure 2.* Sketch example of design ideas by Charles Correa (Lacy 1991)

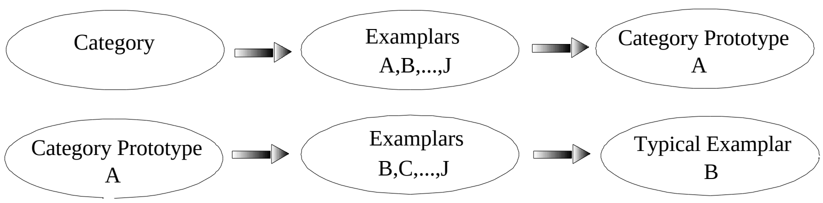


*Figure 3.* Sketch analysis process

## 4. Results and Discussion

### 4.1 RESULTS OF SIMILARITY MEASURE TO SHAPE FEATURE CATEGORIES

Tables 1 and 2 show the similarity measures for “indentation”, “protrusion” (word length 3 ~ 7 Q-codes) and “alternation” (word length 1 ~ 6 Q-codes) shape feature categories. Figures 4 and 5 show graphical representations of each shape’s similarities to these categories.

Table 3 shows the results of similarity measure to five discrete shape feature categories. This is the result of an analysis of the qualitative representation of each shape patterns in terms of types and occurrences of shape features. One of the limitations of this experiment is that this work did not consider the contrasting shape feature category in measuring similarity to a particular category.

TABLE 1. Overall similarity measure for the indentation and protrusion categories

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| Indentation | $t^3s^4$ | $t^2s^3$ | $t^5s^4$ | $t^4s^4$ | $t^1s^4$ | $t^8s^4$ | $t^3s^4$ | $t^2s^4$ | $t^2s^4$ | $t^1s^4$ |
| Protrusion | $t^2s^4$ | $t^1s^4$ | $t^5s^3$ | $t^4s^2$ | $t^0s^5$ | $t^8s^3$ | $t^3s^3$ | $t^2s^3$ | $t^1s^4$ | $t^0s^5$ |
| Sum | $t^5s^8$ | $t^3s^7$ | $t^{10}s^7$ | $t^8s^6$ | $t^1s^9$ | $t^{16}s^7$ | $t^6s^7$ | $t^4s^7$ | $t^3s^8$ | $t^1s^9$ |

TABLE 2. Similarity measure for the alternation category

| Length | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $t^{11}s^1$ | $t^{11}s^1$ | $t^{14}s^2$ | $t^{12}s^2$ | $t^5s^3$ | $t^{20}s^2$ | $t^{10}s^2$ | $t^8s^2$ | $t^{11}s^0$ | $t^5s^3$ |
| 2 | $t^8s^0$ | $t^7s^0$ | $t^{14}s^0$ | $t^{12}s^0$ | $t^4s^2$ | $t^{20}s^0$ | $t^{10}s^0$ | $t^8s^0$ | $t^6s^0$ | $t^4s^2$ |
| 3 | $t^5s^3$ | $t^2s^4$ | $t^{13}s^2$ | $t^{10}s^2$ | $t^3s^4$ | $t^{20}s^1$ | $t^9s^2$ | $t^5s^3$ | $t^2s^4$ | $t^3s^4$ |
| 4 | $t^4s^4$ | $t^0s^6$ | $t^{12}s^3$ | $t^8s^3$ | $t^2s^5$ | $t^{16}s^2$ | $t^8s^3$ | $t^2s^5$ | $t^0s^6$ | $t^2s^5$ |
| 5 | $t^2s^5$ | $t^0s^6$ | $t^{11}s^3$ | $t^6s^3$ | $t^0s^6$ | $t^{20}s^1$ | $t^7s^3$ | $t^0s^6$ | $t^0s^6$ | $t^0s^6$ |
| 6 | $t^0s^7$ | $t^0s^7$ | $t^7s^5$ | $t^4s^5$ | $t^0s^7$ | $t^{20}s^2$ | $t^6s^4$ | $t^0s^7$ | $t^0s^7$ | $t^0s^7$ |
| Sum | $t^{30}s^{20}$ | $t^{20}s^{24}$ | $t^{71}s^{15}$ | $t^{52}s^{15}$ | $t^{14}s^{27}$ | $t^{116}s^8$ | $t^{50}s^{14}$ | $t^{23}s^{23}$ | $t^{19}s^{23}$ | $t^{14}s^{27}$ |

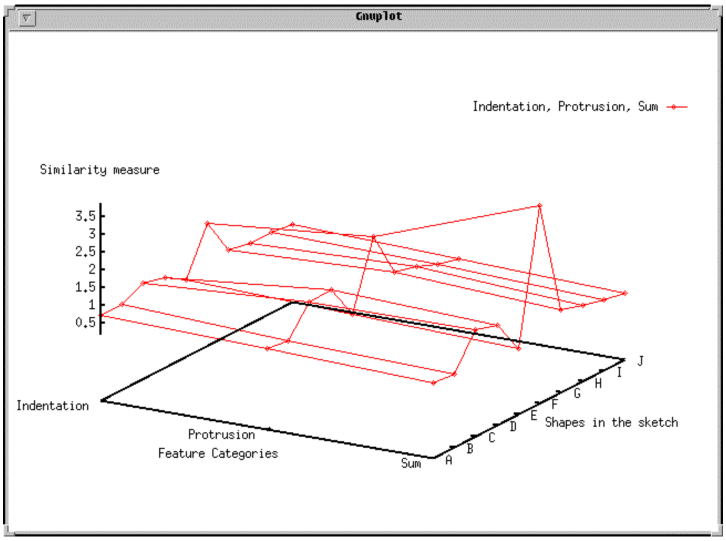


*Figure 4.* Similarity measure for the indentation and protrusion categories (t=1.2, s=0.8)

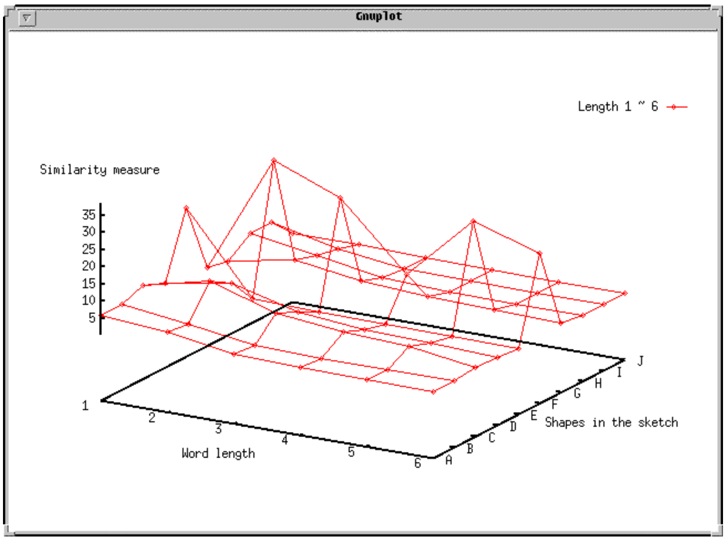


*Figure 5*. Similarity measure for the alternation category (t=1.2, s=0.8)

TABLE 3. Results of similarity measure for the shape feature categories

| Feature Categories | Order of categorical typicality | Categorical prototype | Average example | Least typical |
|---|---|---|---|---|
| Indentation | F-C-D-B-A-G-H-I-E,J | F | A | E, J |
| Protrusion | F-D-C-G-H-A-B-I-E,J | F | H | E, J |
| Alternation | F-C-D-G-A-H-I-B-E,J | F | A | E, J |
| Iteration | F-C-D-G-H-A-B-E,J-I | F | H | I |
| Symmetry | F-C-G-D-A-I-H-B-E,J | F | A | E, J |

## 4.2 DISCUSSION

We have demonstrated shape feature analysis of architectural sketches that takes a qualitative representation of shapes and shape feature categorisation. We have also demonstrated sketch comparisons using similarity measures for the feature category.

This paper clarifies several aspects of conceptual sketch analysis. First, the ambiguous and ill-defined drawing objects can be computationally simulated and compared to give a set of data for the selection of a design alternative satisfying particular criteria. Second, shapes in the sketch level can be treated as a class and as a category in such a way that conceptual design modeling contrasts to final stage design modeling. Third, feature based design provides tools for representation, recognition, comparison and explanation of the sketches.

Conventional computational support for architectural design seems to overlook the significance of possible computer aids for the early stages of designing particularly in the analysis and generation of sketches. Feature analysis of sketches demonstrated in this paper, provides methods and tools to approach architectural sketches in order to explore the early phases of conceptual design.

## Acknowledgements

This research is partly supported by an OPRS scholarship at Sydney University. Computing resources are provided by the Key Centre of Design Computing.

## References

Alexander, C.: 1979, *The Timeless Way of Building*, Oxford Univ. Press, New York.

Coyne, R., Rosenman, M., Radford, A., Balachandran, M., and Gero, J. S.: 1990, *Knowledge-Based Design Systems*, Addison-Wesley, Reading.

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

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

Green, M.: 1992, Conceptions and misconceptions of knowledge aided design, *in* Green, M. (ed), *Knowledge Aided Design*, Academic Press, London, pp. 1-24.

Jared, G. E. M.: 1984, Shape features in geometric modeling, *in* Pickett, M. S. and Boyse, J. W. (eds), *Solid Modeling By Computers: from Theory to Application*, Plenum Press, New York, pp. 121-137.

Lacy, B.: 1991, *100 Contemporary Architects, Drawings & Sketches*, Harry N. Abrams, New York.

Rosch, E and Mervis, C. B.: 1975, Family resemblances: Studies in the internal structure of categories, *Cognitive Psychology*, **7**:573-605.

Rosch, E. and Lloyd, B. B.: 1978, *Cognition and Categorization*, Lawrence Erlbaum, Hillsdale.

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