Recognition of Building Elements in Free-Hand Sketches

Alexander Koutamanis
Faculty of Architecture, Delft University of Technology, The Netherlands
http://caad.bk.tudelft.nl
a.koutamanis@tudelft.nl

Abstract. The paper presents a framework for the recognition of free-hand sketches based on spatial and temporal information that permits identification of strokes, grouping of strokes into graphic symbols and recognition of depictions of building elements in the graphic symbols. It proposes that the symbols involved in 2D projections are generally small and manageable despite the vagueness and multiple layering of sketches.

Keywords: Sketching; recognition; building elements; representation.

Sketches analysed
Despite the attention it has attracted in analyses of architectural creativity, design processes and drawing, sketching has remained a peripheral subject in CAAD. It is frequently seen as an essentially analogue and individual activity, taking place before information enters the computer or in parallel to digital representations (Goldschmidt and Porter, 2004). In recent years affordable sketching tools have stimulated renewed interest in digital sketching for a variety of applications, from drawing education to CAD markup and design collaboration (Cheng and McKelvey, 2005; Koutamanis, 2005). A logical next stage is a new investigation of an old ambition: to transform sketches into CAD drawings and models without losing the vagueness and multiple layering of the original. This transformation involves three main steps:

- Segmentation: subdivision of the sketch into parts tentatively corresponding to meaningful entities. In sketches this normally amounts to recognition of discrete strokes made by sketchers and of clusters formed by these strokes.
- Identification of graphic symbols, i.e. of geometric shapes formed by strokes in the different segments of a sketch. The combination of strokes into discrete objects is essential for the recognition of meaningful entities or relations.
- Recognition of architectural symbols, i.e. depictions of architectural entities by means of graphic symbols, e.g. space bubbles indicating spaces (allocation of activities) or a series of blobs indicating a colonnade.

The present paper focuses on a particular class of architectural symbols, those denoting building elements, such as walls, columns, doors or windows, in 2D projections like floor plans and sections, where solid elements are depicted in section. The appearance of such elements in 3D projections is entirely different (volumetric objects). The choice of focus derives from the analogue character of sketching and the consequent emphasis on the ‘solids’ of a design, i.e. building elements. Most graphic symbols are either direct depictions of building elements or describe conditions directly related to aspects of building elements like spatial arrangement. As a result, the transition from graphic symbols to symbols of building elements is assumed to be relatively short and less ambiguous than with spaces.
The parts of a sketch

Transformation into a CAD document concerns the following types of sketch primitives (Koutamanis, 2001):
1. Solid lines, usually drawn with a single stroke
2. Multiple lines, drawn with many strokes at approximately the same location; multiple lines are not necessarily adjacent in the sketching sequence
3. Dashed lines, drawn with a series of collinear short strokes
4. Blobs: small, compact clusters focused on a crucial point or indicating an element with a small footprint, e.g. an intersection of grid lines or a column

- Textures: repetitive patterns usually characterized by translational symmetry and drawn with short strokes
- Hatches: a particular class of textures, usually drawn with continuous sub-sequences of parallel solid or dashed lines
- Grids: from a graphic point of view grids can be considered as hatches but semantically they indicate coordinating schemata (e.g. for the positioning of load-bearing elements) rather than fill surfaces. This is reflected in the scale (spacing) of grid elements.

These primitives combine on the basis of a limited number of relations to form more complex entities (e.g. dashed lines or shapes like crosses and squares) as well as patterns indicating or occupying 2D surfaces in the sketch, such as:
Graphic symbols

Identification of graphic symbols is based on a number of working hypotheses:

1. Symbols are formed by combining strokes of the same type. Composite symbols are produced by the combination of two or more discrete symbols.
2. While it is often due to mechanical constraints, the temporal order of strokes reveals a clustering of actions that relates to both graphic and architectural symbols: repetitive parts in the sketching sequence are generally confined to a single class or group of symbols.
3. Each stroke is part of a single graphic symbol, resulting into a tree structure at the lower levels of the recognition process.

The process starts with the vectorization of strokes, which is performed externally in the input media used (digital pen, graphics tablet or smart-board). Low-level graphic primitives then categorized as either solid lines or blobs on the basis of spatial criteria (length, compactness and direction of stroke). Recognition of complex symbols and patterns takes place in two subsequent stages by grouping graphic entities together on the basis of predefined and learned templates. The predefined templates recognize the following graphical symbols: multiple lines, dashed lines, L-shapes, T-shapes, H-shapes, crosses, open and closed polygons, blobs and asterisks.

These templates accommodate both spatial and temporal characteristics (Gross, 1996), organized in a bottom-up fashion as flexible groups of strokes (Mahoney and Fromherz, 2002). The bottom-up definition makes templates adjustable as to the number of strokes. For example, a second stroke for emphasis or because of sketching errors can be neglected or fused with the first one. The inclusion of so-called don't care elements in the template increases flexibility (Persoon, 1988; Sharon and Van de Panne, 2006) and makes recognition insensitive to small gaps and perturbations that may survive stroke normalization.

Learned templates are acquired in each session in a similar bottom-up fashion. These templates make more intensive use of temporal information because of the close relations between repetition and mechanical constraints. They are triggered by existence of repetition in the sketching sequence and perform the following functions:

1. Identify segments of a sequence (adjacent strokes) as symbols
2. Correlate two or more stroke sub-sequences on the basis of common spatial aspects between corresponding strokes
3. Create templates for symbols encountered in the patterns
4. Create pattern templates on the basis of their constituent symbols and repetition characteristics (instead of precise metrics of sub-sequences)

Pattern templates include tolerances in the definition of the constituent symbols (redundant strokes that duplicate earlier ones, especially if mechanically unsuccessful, as well as missing strokes –a common occurrence in repetitive actions) and in the spatial arrangement of these symbols (e.g. multiple grids).

Building elements

Recognition of building elements is performed by similar templates which identify architectural symbols in single graphic symbols or in spatially related groups of graphic symbols. Recognition results are often tentative and sometimes ambiguous, involving multiple interpretations. Our current solution to such problems is to accept all probable interpretations and indicate all possibilities by placing interconnected clones of the graphic symbols on all appropriate CAD layers. Once the user has confirmed one interpretation (usually by a relevant action on a particular layer such as a translation and alignment to other objects) all other clones are discarded.

Building elements are encountered at two main levels of abstraction:
1. Discrete, specific building elements, e.g. a blob representing a column or a rectangle consisting of four solid lines representing a wall.

2. Abstract elements or clusters, e.g. a dashed line indicating a colonnade or the axis of a cluster of load-bearing elements.

The distinction between these two levels should not be confused with architectural abstraction (e.g. representing a wall by a single solid line) and symbolization (e.g. a short line perpendicularly drawn on a wall symbol as indication of the presence of a door). Such abstract symbols are treated as members of the first level.

The criteria used in practically all templates are scale, position and line type. Their significance and priority varies with the type of the architectural drawing. Scale refers both to the size of a symbol relative to others in the sketch and to the relations between components of a symbol. For example, an H-form is a common symbol for a window (and in some notations of a door) but only if the two end line segments are much shorter than the connecting line segment (roughly 1:5 or lower in the cases studied). However, the main criterion for this particular form is its positioning in a wall (the purpose of the two end line segments being to accentuate the opening).

Figure 3
Schematic outline of the recognition process
Interestingly, the connecting line segment may be absent. This construction is common to all kinds of openings: position is generally the main criterion but mainly for graphic symbols that qualify by scale.

**Spatial and temporal issues**

Recognition of symbols refers primarily to spatial information, i.e. the form and location of strokes – the paradigmatic dimension of a drawing (Van Sommers, 1984). The syntagmatic dimension, i.e. the temporal sequence of sketching actions improves the effectiveness and reliability of information by facilitating the distinction between intentional and accidental spatial relationships (Koutamanis, 2005). A sketcher often enters a symbol (especially a simple one) with temporary adjacent strokes. The most notable exception is the use two or more sub-sequences to enter a pattern of symbols.

The obvious reason for doing so is sketching ease: the mechanical constraints of almost all input media make even the novice sketcher enter strokes of a particular orientation sequentially before moving on to related strokes with a different orientation. This applies to even relatively simple forms such as rectangles: many sketchers enter the vertical sides first and the horizontal ones afterwards. This means that the recognition system should evaluate the spatial evidence for a symbol on the basis of temporal information, e.g. accept a cross if it is formed of perpendicularly intersecting strokes $m + i$ and $n + i$, $m$ and $n$ being the starting strokes of respectively the horizontal and vertical sub-sequences.

The most challenging forms are produced by several sketchers or with strokes of different types. Most of these forms fall under three main categories:

1. **Completion**: the shape is formed by strokes entered by several sketchers who typically copy the style of earlier strokes and align their own strokes to them.
2. **Emergence**: a shape is produced unintentionally, frequently through variation or emphasis. Such shapes are classed as suspect (i.e. accidental products of the multiple layers of a sketch) but not discarded because of the impact they can have on the re-interpretation of the sketch (Gross, 2001; Koutamanis, 2005).
3. **Loose correlation**: a shape comprises strokes of different types that may be dispersed in the sketching sequence and may have been made by various sketchers. In most such cases the shape is intentional and the differences between strokes indicate different properties, e.g. material or performance differentiation between bounding elements.

In contrast to spatial information, the syntagmatic dimension of a sketch does not provide strong cues for the recognition of a shape. The main reason is that a shape can be produced by different stroke sequences resulting into equivalent shapes. This is apparent in multiple strokes, where overtracing is used not merely to edit an existing line but for a variety of additional purposes, including fuzzification and emphasis. A useful syntagmatic cue in multiple strokes is the use of a single stroke as organizational line before the other strokes are made, usually for all sides of a symbol. These initial strokes serve as an initial spatial definition as well as an anchor for the sketcher’s interaction with the sketching media (mechanical dimension).

Use of syntagmatic information for the evaluation of preliminary recognition results from spatial cues presupposes:

- Identification of homogeneous sub-sequences by segmenting of the sketching sequence on the basis of stroke similarity and repetition
- Correlation of sub-sequences by the same sketcher on the basis of length and common starting area
- Correlation of sub-sequences and strokes by different sketchers so as to identify cases of over-tracing and completion

These make it possible to identify symbols drawn with sequential strokes as well as with non-sequential ones but following consistent spatial and temporal patterns. Interpretation of the resulting graphic
elements into architectural symbols is based on spatial information (scale and connectivity). The same process applies to the recognition of patterns such as a linear arrangement of blobs and its interpretation as a colonnade. The most interesting aspect of pattern recognition is that they are normally entered by a single sketcher in a single sub-sequence or a few sequential sub-sequences. Modification of a pattern (extension, translation, rotation or scaling) may involve other sketchers too who usually copy the style of the initial sketcher.

Simple patterns like hatches are evident in the syntagmatic sequence but distinction between different types relies heavily on spatial information, in particular scale. For example, both grids and hatches are drawn with similar strokes (usually thin solid lines) and similar repetition patterns. Their essential difference lies in the spacing of strokes. In the cases we have been studying this distance in hatches was up to 4 mm while in grids it was significantly greater than 4 mm.

Discussion: performance and further development

The performance of the system has been evaluated in eight simulated design sessions. The products of these sessions were drawings like those in Figure 4.

From a technical viewpoint 73% of the strokes in these sketches were recognized as being part of a graphic symbol. 11% of the remainder were lone strokes. The recognition results could have been positively influenced by the self-consciousness of users in the simulated design sessions: even though they received no instructions and just a minimal training with the digital pens or graphics tablets, they tended to be much neater in their drawing than one would have expected.

The small number of symbols recognized in these sessions agrees with the findings of earlier research (Do and Gross, 1997). Interestingly some of the symbols were clear products of common authorship, proposed by one sketcher and confirmed or elaborated by others, resulting into a uniform composite graphic symbol denoting a particular class of architectural entities. Such occurrences were relatively common in the sessions we analysed. The development of a common symbol combining strokes by various sketchers appeared to imply some kind of agreement concerning the character of the depicted architectural elements, as well as the acceptance and priority of different constraints and properties.

Probably the most striking result from these sessions was that recognition of symbols relied less on learning than initially assumed. Identification of most symbols involved few strokes and even fewer relationships between strokes. Having said that, the system has still many limitations: it is yet incapable of identifying relationships between different symbols and notations and the complexity of graphic

Figure 4 Sketches from simulated design sessions
symbols is arguably much lower than that of architectural symbols. It is therefore possible that learning is more significant at that level. With respect to graphic symbols, the complexity of many sketches turns out to relate to fuzziness (e.g. use of multiple strokes to indicate uncertainty about size or location) and the multi-layered character of sketches which cannot be unraveled without reliable syntagmatic information.

The main problems in recognition were the uncertainty concerning alternative versions of a building element, especially if drawn by different sketchers, and a small number of illogical relations between building elements, e.g. a free-standing door or a wall blocking a corridor. This relates to the two main limitations of the current version of the system: its inability to operate in near real time (arguably due to the lack of meta-learning but also because of the asynchronous integration of the digital pens –our preferred input media– into the computer environment where processing takes place) and its lack of self-sufficiency in the recognition of contextual constraints, for which we still need video evidence (Mahoney and Fromherz, 2002).

While these may relate too closely to our approach and hypotheses, there is room for improvement in a number of aspects. Recognition of annotations is probably the highest priority. This involves higher selectivity, i.e. more attention for salient details like arrowheads, not only for the distinction between annotations and other graphic symbols but also for the identification of the elements or relationship to which the annotation refers. This is of particular interest also for the recognition of graphic elements because transformations of existing symbols (e.g. translation of a grid or rotation of a colonnade) are frequently indicated by diagrammatic means. Recognition of annotations also involves an increase in uncertainty and ambiguity because some graphic elements that assume different roles in the same sketch. A hatch, for instance, can be used both as graphic annotation of a space and as definition of an area.
References


