RECOGNITION AND INDEXING OF ARCHITECTURAL FEATURES IN FLOOR PLANS ON THE INTERNET

ALEXANDER KOUTAMANIS
Faculty of Architecture, Delft University of Technology
Berlageweg 1, NL-2628 CR Delft, The Netherlands
Email: a.koutamanis@bk.tudelft.nl

Abstract. The Internet promises a worldwide information system, capable of uniting different sources and types of original, up-to-date and directly usable information. Among the main components of this system are retrieval mechanisms characterized by high precision and recall, as well as by supportive relevance feedback. The textual versions of these retrieval mechanisms have been available for some time and have achieved a certain degree of efficiency and sophistication. Image retrieval lags behind, despite the recent advances in content-based retrieval. In architecture this is largely due to the lack of integration of domain knowledge and known formalisms. Indexing and retrieval of architectural floor plans can rely on existing generative systems such as shape grammars and rectangular arrangements. By reversing generative systems in purpose we derive compact descriptions that describe completely a floor plan and make explicit all relevant features rather than a small number of features. The main limitation of reversed generative systems is that they apply to specific classes of designs. Unification in indexing and retrieval can only take place at the level of basic primitives, i.e. spaces and building elements. In both vector and pixel images of architectural floor plans this can be achieved by a universal recognition system that identifies salient local features to produce a basic spatial representation.

1. Indexing and Retrieval in a Worldwide Information System

As a result of ongoing rapid technological and social changes, we have entered a period characterized by an unprecedented democratization of information and computing technologies. The ability to generate, disseminate and retrieve information is increasingly within the reach of a large number of people who are less bounded by constraints traditionally associated with information channels such as periodicity and censorship. The main carrier of such activities, the Internet, is portrayed as a worldwide system that will ultimately unite different sources and types of information into a ubiquitous infrastructure. What makes this infrastructure particularly appealing is that it comprises original, up-to-date information. Rather than evaluating the quality of information by the authority of the channel we are promised the possibility to evaluate the reliability and quality of the actual source (Koutamanis, 1998). Moreover, we are offered ample opportunity to use this information directly and unobtrusively, e.g. by integrating the latest building components from an online database or by evaluating the conformity of a design to building codes and regulations (Koutamanis, 1997).

Straightforward production and efficient dissemination of information are the hallmarks of the Internet today, especially in popular areas such as entertainment. In technical and professional areas like architecture and building development is less spectacular but nevertheless sufficient as an indication of the future. Publication of design and product documentation on the Internet is available through most drawing and modeling programs but the purpose of publication generally does not go beyond communication. In particular visual documentation is frequently treated as mere illustration.

The explosive growth of the Internet and its inherent lack of quality control may offer possibilities to the producer and publisher but may also make retrieval of information time-consuming, tedious, unreliable and unproductive. This problem has been attacked in two different ways. The first is the use of expert or just informed *search intermediaries*. These either operate as domain authorities who sanction information sources or as peer judges who guide search through pre-selection (Negroponte, 1995). The second way is through search engines that index Internet sites practically exhaustively and facilitate full text / natural language retrieval. A combination of the two is based on vocabulary control, i.e. the classification of indexing terms into thesauri (GAHIP, 1990). These allow for non-redundant, coherent and consistent indexing and provide knowledgeable relevance feedback already from the stage of query formulation (Koutamanis, 1995). At the same time they avoid the pitfalls of authoritative control of the actual information and its sources and focus instead on the improvement of information utility.

Retrieval mechanisms are already an integral part of the Internet. Link lists and search engines are routinely used for retrieving information. Provided the query is successfully formulated, *recall* (the ratio between the number of relevant documents retrieved and the number of all relevant documents in the system) is generally high, despite the usually frustration of obsolete links. On the other hand, *precision* (the ratio between the number of relevant documents retrieved and all documents retrieved by a specific query) is lowered by the lack of clear terminologies and poor contextual matching but remains mostly adequate. Textual search engines are increasingly refining their query strategies, e.g. by integrating elements of vocabulary control, improving indexing efficiency and providing relevance feedback.

Consistently with the general downplaying of visual information to the level of illustration in text documents, retrieval of images lags behind text retrieval. In information generation and indexing images are treated as empty vessels, scarcely ever annotated with textual indexing terms. Consequently, image retrieval was until recently largely ignored as an imprecise, vague area. The current attention on *content-based image retrieval* is doing much to redress the lack of attention to imagery as information. Using image-processing and pattern-recognition techniques content-based approaches derive features from low-level attributes such as color, texture and shape directly from the image. These features form the basis for computing image distance measures (i.e. similarity) between user input and the images in a database.

In terms of efficiency and reliability content-based approaches appear to offer significant advantages over standard databases that depend on keywords that describe image features. Indexing an image database with keywords requires large numbers trained human classifiers and unambiguous but flexible keyword conventions. However, content-based systems may also require extensive human effort in e.g. manual image segmentation (as in QBIC) or annotation (in Chabot). Another advantage is the ability to form queries through visual interfaces, e.g. by selecting from canonical or typical images in order to retrieve the images of a particular category.

A severe limitation of many content-based systems in their application to architectural drawings is reliance on global, low-level features such as color. Perceptual similarity between two photographic images could rely on such features but in architecture shape is generally more important, as it relates to typical architectural queries concerning type, style and spatial arrangement. This presupposes automated image segmentation, on the basis of not only color (as in VisualSEEK or Photobook) but also shape. This recovers tentative objects which combine to form the image. The information in the image is largely dependant upon these objects and their configuration.

Analysis of vector or pixel images can recover important features, such as the overall shape of a floor plan or a space. For practical reasons, however, image analysis and content-based retrieval are generally restricted to a small number of salient features (minimization of keywords). These form the basic classification criteria for the image in the framework of specific indexing and retrieval activities. Therefore, content-based image indexing becomes a deterministic search for the most economical collection of features (Gong, 1998). Even when the objective is just the vectorization of a pixel image, efficiency dictates that issues such as drawing style are ignored because they are deemed secondary in many searches (Baird, Bunke and Yamamoto, 1992).

A small number of indexing features implies a well-defined, probably narrow application area. An architectural historian has arguably other interests and consequently uses different query terms than a designer in search of precedents or a real-estate advisor who must accommodate an organization in a stock of available buildings (Leusen and Mitossi, 1998). One possibility is to devise distinct indexing and retrieval systems for each application. However, these applications and the corresponding indexing terms overlap to a degree that suggests the possibility and utility of systems that cover at least a wide area of applications on the basis of a coherent collection of design aspects rather than opportunistic definitions of a minimal feature set. Spatial characteristics are one such typical collection that relate to a wide range of applications and a correspondingly large number of indexing / query terms.

2. Architectural Floor Plans as Carriers of Spatial Design Information

Content-based systems appear to focus on images such as photographs and video. These represent a large percentage of architectural heritage but from a design viewpoint are arguably less significant than drawings (Evans, 1995). Drawings remain the core of design documentation and contain spatial and building information of particular value to precedent and case-based approaches. For this reason, drawings and especially orthographic projections such as floor plans form a priority in the dissemination of architectural information on the Internet.

Floor plans have proved an effective representation of spatial characteristics that combines ready perception of formal aspects with measurability. Architectural documentation includes a large number of floor plans that describe the spatial articulation of important designs with clarity and succinctness. These images usually serve as the basis for analysis and presentation. Floor plans also form a recognition challenge in that they deviate from the images that have been the main target of content-based retrieval. The main deviation lies in the conventional / geometric structure of floor plans that makes them legible only after one has been accustomed to the conventions of the representation (Lopes, 1996).

It is noteworthy that floor plans may have a more promising future than other types of architectural drawings. Computerization has made three-dimensional modeling of buildings efficient, affordable and popular. Interactive three-dimensional models represent a preferable source of spatial and building information that combines intuitive understanding with flexibility and measurability. Elevations and even sections are increasingly being replaced or integrated in three-dimensional models. Floor plans, on the other hand, are either treated as the basis of the three-dimensional models or as a complementary basic representation at a higher abstraction level for more analytical purposes.

3. Indexing Through Generative Systems: Shape Codes

Comprehensive indexing and retrieval of floor plans can make use of existing generative systems with underlying representation formalisms, such as shape grammars (Stiny, 1975; Stiny, 1980) and rectangular arrangements (Steadman, 1976; Steadman, 1983). A well-defined generative system produces a consistent and comprehensive collection of patterns that meet certain constraints, usually those of a particular style. The reversal of a generative system returns *shape codes:* compact descriptions (strings) that describe completely a floor plan and make explicit all relevant features rather than the small number of features used in content-based image retrieval. For example, in the prototypical Palladian grammar (Stiny and Gips, 1978; Stiny and Mitchell, 1978; Stiny and Mitchell, 1978), spatial articulation is produced by the following collection of shape rules:

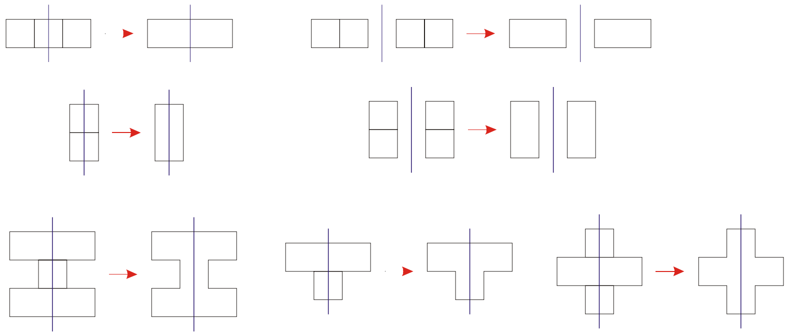

Figure 1. The Palladian shape rules: $R = \{R_{12}, R_{13}, R_{14}, R_{15}, R_{16}, R_{17}, R_{18}\}$ (left to right, top to bottom)

These room layout rules are applied recursively to a 3 x 5 grid until a floor plan layout is generated. This grid, the acknowledged reference framework of Palladian villas (Wittkower, 1952), serves as the initial shape (I_w) for the Palladian floor plans. The shape codes for the families of the eight villa plans actually by Palladio in the catalogue of the Palladian grammar are:

Emo: $S_{11} = I_w\ R_{12}$
Ragona: $S_{49} = I_w\ R_{12}\ R_{13}$
Zeno: $S_{53} = I_w\ R_{12}\ R_{14}\ R_{13}$
Badoer & Poiana: $S_{56} = Iw\ R_{12}\ R_{14}\ R_{15}$
Pisani: $S_{120} = Iw\ R_{12}\ R_{16}$
Sarraceno: $S_{135} = I_w\ R_{12}\ R_{17}\ R_{15}$
Malcontenta: $S_{204} = I_w\ R_{12}\ R_{18}\ R_{13}$

Indexing of the floor plans by their shape code affords comprehensiveness and flexibility. For example, the central cross-shaped space in Villa Malcontenta appears as R_{18} in its shape code and elongated spaces in the lateral flanks are denoted by R_{15}. Also significant is the *absence* of certain rules from the shape code. A shape code without R_{16} or R_{18} means that there is no unified central space, as in Villa Emo. Absence of R_{13} and R_{15} reveals that there has been no enlargement of spaces in the lateral flanks, as in Villa Emo and Villa Pisani.

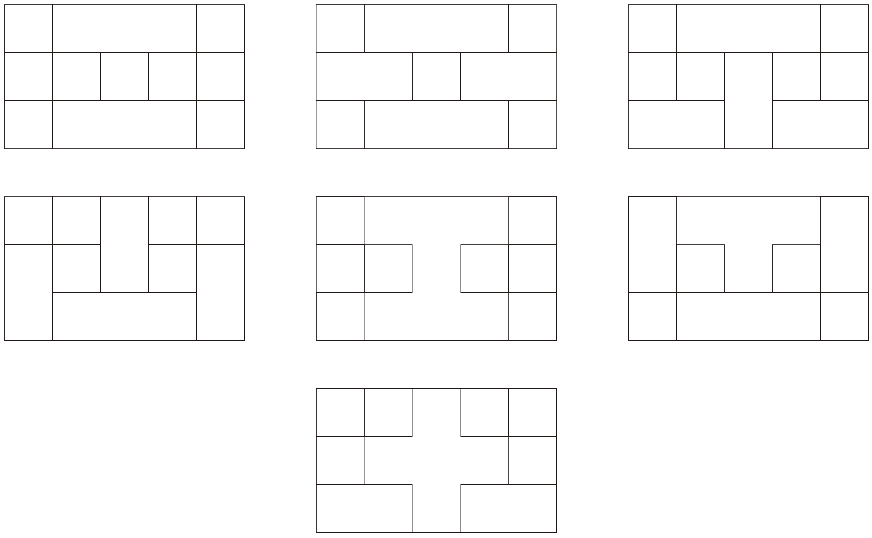

Figure 2. The Palladian villas: Emo, Ragona, Zeno, Badoer / Poiana, Pisani, Saraceno and Malcontenta (left to right, top to bottom)

This primary shape code can be extended to indicate the number of times a rule is applied. This permits identification of the number of spaces with a particular shape and orientation, as well as direct calculation of the precise number of spaces in the floor plan. For each application of a particular rule a specific number of rooms should be deducted from the fifteen of I_w.

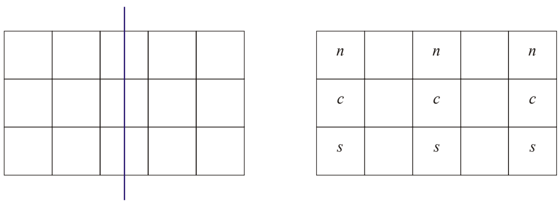

Figure 3. Localization scheme for I_w

A further elaboration of the shape code is *localization* of rules so as to permit precise identification of features with respect to the initial shape I_w. This is achieved by specifying the subshape of I_w to which a rule is applied, e.g. on the basis of a north-center-south orientation scheme in the initial shape. In this scheme the cells belonging to the northern, central and southern rows are labeled respectively with *n, c* and *s.* The intermediate columns (second and fourth) are not labeled. The addition of localization means that a pattern can be fully reconstructed from its shape code. In addition, floor plans become oriented. In the Palladian example, localization makes patterns oriented in the north-south (vertical) direction. For example, the patterns $I_w\, R_{12}(n)\, R_{12}(s)$ and $I_w\, R_{12}(s)\, R_{12}(n)$ are not identical but symmetrical. Left and right remain interchangeable, in accordance with the bilateral symmetry of Palladian floor plans.

A final elaboration is compression of the shape code. Certain rules presuppose and subsume other rules. For example, R_{12} is a prerequisite to R_{16}, R_{17}, and R_{18}. The horizontally elongated space produced by R_{12} becomes part of the complex shapes returned by these three rules. Consequently we can remove such instances of R_{12}. The oriented shape code can still reproduce the complete pattern, as e.g. $R_{18}(c)$ in the shape code of Villa Malcontenta becomes shorthand for $R_{12}(c)\ R_{18}(c)$ and $R_{17}(n)$ in Villa Sarraceno implies $R_{12}(n)\ R_{17}(n)$. Compression returns the following shape codes which describe the corresponding plans uniquely and unambiguously:

Emo: $S_{11} = I_w\, R_{12}(n)\, R_{12}(s)$
Ragona: $S_{49} = I_w\, R_{12}(n)\, R_{12}(s)\, R_{13}(c)$
Zeno: $S_{53} = I_w\, R_{12}(n)\, R_{14}(s)\, R_{13}(s)$
Badoer & Poiana: $S_{56} = I_w\, R_{12}(s)\, R_{14}(n)\, R_{15}(s)$
Pisani: $S_{120} = I_w\, R_{16}(c)$
Sarraceno: $S_{135} = I_w\, R_{12}(s)\, R_{17}(n)\, R_{15}(n)$
Malcontenta: $S_{204} = I_w\, R_{18}(c)\, R_{13}(s)$

4. Applicability and Extensions

Reversed generative systems facilitate indexing of floor plan images through the registration of the generative process in shape codes. Queries can be made in terms of explicit or implicit parts of a shape code, from the number of spaces in the plan and the shape of specific spaces to the geometric and topological arrangement of spaces, including relations such as symmetry and collinearity, adjacency and access. The sum of these relates to the conformity of the floor plan to the constraints of the system, i.e. its quality with respect to the corresponding style (Stiny and Gips, 1978; Koutamanis, 1997). This opens the way to more generic queries, such as “Palladian designs” or “classical floor plans”.

The shape code is related to theories of perception and recognition, in particular the *recognition-by-components* (RBC) theory and the *structural information theory* (SIT). According to RBC a three-dimensional image can be reduced to a small repertory of twenty-four basic components, called *geons*, which are applicable to all kinds of scenes and images (Biederman, 1987;

Biederman, 1995). Geons could replace spaces and abstract building elements when shape coding is applied to three dimensional objects, so as to link perception of built form with architectural knowledge. SIT is important for the processing of shape code. In SIT the primary code of a pattern is minimized by repeatedly and progressively transforming the primitive code on the basis of iteration, reversal, symmetry, distribution and continuation. This process returns an end code, whose structural information cannot be reduced further (Leeuwenberg, 1967; Leeuwenberg, 1971). Therefore, SIT provides the means for abstracting the shape code without loss of schemata such as repetition and symmetry. In this way, we can derive the different descriptions an image affords and choose the one that contains the least information, i.e. satisfies the constraints of a particular formal context.

The main disadvantage of reversed generative systems is that they apply to specific classes of designs. As research into shape grammars and syntactic pattern recognition has demonstrated, a universal generative system that can produce all possible classes seems unattainable (Fu, 1974; Fu, 1982). Consequently, unification of the reversed versions for indexing and retrieval can only take place at the level of basic primitives, i.e. spaces and building elements. However, these are frequently only implicit in the generative system, as e.g. the nine spaces in I_w.

5. Recognition of Spaces and Building Elements

Consistency among different reversed generative systems presupposes a universal basis, such as a basic spatial representation of floor plans in terms of their spaces and building elements. Automated indexing and retrieval of these elements can be achieved by a universal recognition system that recovers these primitives in all kinds and styles of floor plans. Such a system should also facilitate the transformation of pixel (scanned) and vector (CAD) drawings into symbolic representations of floor plans. These representations form the input of higher indexing systems such as a reversed shape grammar. Such a transformation goes beyond the capabilities of most vectorization systems, which fail to address the semantic content and structure of a floor plan.

As with the indexing of floor plans, recognition of spatial and building primitives relates to domain knowledge. Rather than relying on low-level geometric information, as with edge-following techniques, we could focus on salient local features that constrain our perception and understanding of architectural representations. In rectangular floor plans, for example, spaces can be recognized on the basis of a typology of just eight space corners (Figure 4). Each corner type is characterized by concrete connectivity expectations in both directions. By propagating recursively these expectations from each corner in an image, we derive feature loops that correspond to the outline of each space. The use of such features permits a higher level of abstraction that is not hampered by imperfections in the image or insignificant properties such as small perturbations in a wall (Koutamanis and Mitossi, 1992; Koutamanis and Mitossi, 1993). This approach is applicable to the recognition of pixel and vector images, with respect to both spaces and building elements (Koutamanis and Mitossi, 1993; Koutamanis, 1995).

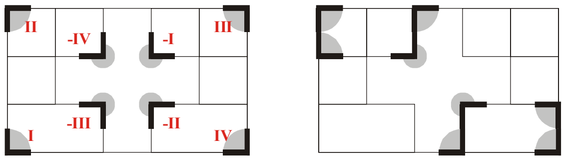

Figure 4. Rectangular space corner typology (left) and connectivity example (right)

6. Summary

Indexing and retrieval of architectural drawings goes beyond the global features and manual annotations with textual terms that characterize most current content-based systems. Relevant queries focus normally on local features and aspects that derive from the symbolic, conventional and semantic dimensions of architectural representations. The derivation and use of such features relates to two different levels of interpretation. The first concerns overall spatial structure that can be derived from existing representational formalisms and corresponding generative systems. The main limitation of this level is that unification of different systems is possible only in terms of basic primitives, i.e. a symbolic representation in terms of spaces and building elements. This representation is the subject of the second level, where relevant features in architectural drawings are recognized and processed towards a general-purpose description that can be used as input for the reversed generative systems.

References

Baird, H. S., H. Bunke, and K. Yamamoto (eds.): 1992, *Structured document image analysis.* Berlin, Springer-Verlag.

Biederman, I.: 1987, “Recognition-by-components: A theory of human image understanding.” *Psychological Review* **94**(2): 115–147.

Biederman, I.: 1995, Visual object recognition. *Visual congition. An invitation to cognitive science. 2nd ed.* S. M. Kosslyn and D. N. Osherson (eds.). Cambridge, Mass., MIT Press.

Evans, R.: 1995. *The projective cast. Architecture and its three geometries.* Cambridge, Mass., MIT Press.

Fu, K.S.: 1974, *Syntactical methods in pattern recognition.* New York, Academic Press.

Fu, K.S.: 1982, *Syntactical pattern recognition and applications.* Englewood Cliffs, N.J., Prentice-Hall.

G.A.H.I.P.: 1990, *Art and architecture thesaurus.* New York, Oxford University Press.

Gong, Y.: 1998, *Intelligent image databases.* Boston, Kluwer.

Koutamanis, A.: 1995, Recognition and retrieval in visual architectural databases. *Visual databases in architecture. Recent advances in design and decision making.* A. Koutamanis, H. Timmermans and I. Vermeulen (eds.). Aldershot, Avebury.

Koutamanis, A.: 1997, Multilevel representation of architectural designs. *Design and the net.* R. Coyne, M. Ramscar, J. Lee and K. Zreik (eds.). Paris, Europia Productions.

Koutamanis, A.: 1997, On the evaluation of figural architectural goodness: A foundation for computational architectural aesthetics. *CAAD Futures 1997.* R. Junge, (ed.). Dordrecht, Kluwer.

Koutamanis, A.: 1998, Information systems and the Internet: towards a news counter-revolution? *4th Design and Decision Support Systems in Architecture and Urban Planning Conference.* Eindhoven.

Koutamanis, A. and V. Mitossi: 1992, Automated recognition of architectural drawings. *Proceedings of the 11th International Conference on Pattern Recognition.* Los Alamitos, IEEE Computer Society Press.

Koutamanis, A. and V. Mitossi: 1993, Adding visual recognition to the capabilities of computer-aided design. *Visualization and intelligent design in engineering and architecture.* J. J. Connor, S. Hernandez, T.K.S. Murthy and H. Power (eds.). London, Elsevier.

Koutamanis, A. and V. Mitossi: 1993, "Computer vision in architectural design." *Design Studies* **14**(1): 40–57.

Leeuwenberg, E.L.J.: 1967, *Structural information of visual patterns. An efficient coding system in perception. Doctoral dissertation, Catholic University of Nijmegen.* The Hague, Mouton.

Leeuwenberg, E.L.J.: 1971, "A perceptual coding language for visual and auditory patterns." *American Journal of Psychology* **84**: 307-350.

Leusen, M. v. and V. Mitossi: 1998, A practical experiment in representation and analysis of buildings. *4th Design and Decision Support Systems in Architecture and Urban Planning Conference.* Eindhoven.

Lopes, D.: 1996, *Understanding pictures.* Oxford, Clarendon.

Negroponte, N.: 1995, *Being digital.* London, Hodder & Stoughton.

Steadman, J.P.: 1976, Graph-theoretic representation of architectural arrangement. *The architecture of form.* L.J. March (ed.). Cambridge, Cambridge University Press.

Steadman, J.P.: 1983, *Architectural morphology.* London, Pion.

Stiny, G.: 1975, *Pictorial and formal aspects of shape and shape grammars.* Basel, Birkhäuser.

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

Stiny, G. and J. Gips: 1978, *Algorithmic aesthetics. Computer models for criticism and design in the arts.* Berkeley, California, University of California Press.

Stiny, G. and J. Gips: 1978, "An evaluation of Palladian plans." *Environment and Planning B* **5**: 199-206.

Stiny, G. and W.J. Mitchell: 1978, "Counting Palladian plans." *Environment and Planning B* **5**: 189-198.

Stiny, G. and W.J. Mitchell: 1978, "The Palladian grammar." *Environment and Planning B* **5**: 5–18.

Wittkower, R.: 1952, *Architectural principles in the age of humanism.* London, Alec Tiranti.