Rough Document Interpretation by Perceptual Organization

Eric Saund, David Fleet, James V. Mahoney, and Daniel Larner
Perceptual Document Analysis Area
Palo Alto Research Center
3333 Coyote Hill Rd. Palo Alto, CA 94304
{saund, fleet, jvmahon, larner}@parc.com

Abstract

Not all document images of interest are comprised of properly scanned, cleanly printed text in a known language, formatted according to standard layout conventions. We are often interested in the content of rough documents. Rough documents include handwritten notes, sketches, drawings, annotations, doodles, specialized notations, unconventional layouts, and poorly imaged markings. Despite their unruly diversity, rough documents embody visual structure which is accessible and exploitable by humans by virtue of our powerful perceptual apparatus. We believe that to achieve breadth and depth in interpreting images of rough documents, computer systems will require a foundation of image analysis approximating the human visual stage postulated as Perceptual Organization. This whitepaper motivates and outlines a research program along these lines. A viewpoint that raises the identification of visual perceptual structure as a primary objective helps to open a broader range of tasks for document image analysis, beyond character-to-text transcription.

1 Introduction: Motivation

1.1 Rough Documents

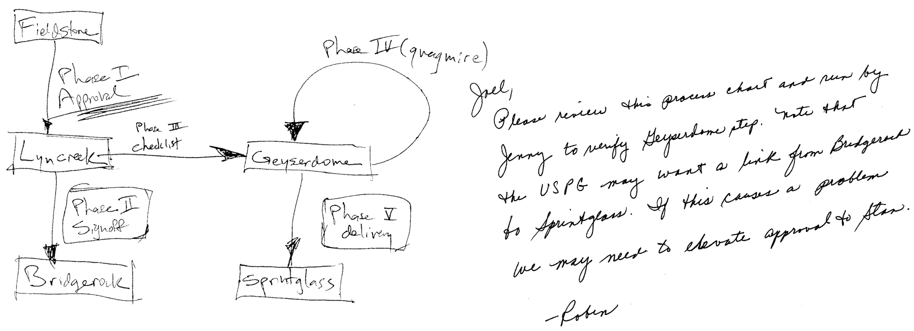
To a human observer, Figures 1a and 1b depict more-or-less the same information. While current text and graphics document recognition technology is capable of transcribing 1b into ascii text and graphical shape models, 1a is far out of reach. Yet, many applications are more likely to encounter images resembling Figure 1a, that is, arising from unpredictable sources and uncontrolled imaging conditions. We may refer to documents whose image content extends beyond properly scanned, cleanly printed text in a known language, formatted according to standard layout conventions, as *rough documents*. Rough documents include handwritten notes, sketches, drawings, annotations, doodles, specialized notations, unconventional layouts, and poorly imaged markings. Figure 2 presents more examples of rough documents whose interpretation by machines is far beyond current technology.

The difficulties posed by rough documents are not just a matter of image degradation due to fading ink, multiple-generation copies, or poor quality scanning. Certainly imaging quality can be a factor, as shown in Figures 2a, 2b, and 2f, which were probably captured with digital cameras. But more significantly, rough documents are typically characterized by unconstrained semantic content rendered in idiosyncratic styles, by casual and imprecise marking processes. The result is tremendous unpredictability in what textual and graphical material needs to be recognized, and tremendous variability in how this material appears in a document image.

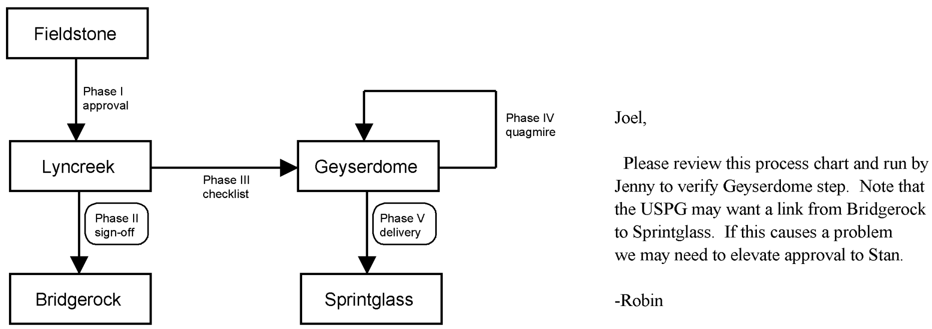
Despite their unruly diversity, rough documents nonetheless embody visual structure which is accessible and exploitable by humans by virtue of our powerful perceptual apparatus. At a glance we identify not only lines and columns of written material, but also printed or hand drawn underlines, encirclings, arrows, figures, the distinct slants of annotations, notable features arising from logos or rubber stamps, coffee stains, graphical arrangements, and so forth. Content elements of this sort are found across document domains. They are all ingredients of comprehensive interpretations of images, and they contribute to critical base-level representations of the layout and coarse-level content of a document necessary to orient and direct localized OCR/ICR, symbol classification, and graphics recognition procedures. We believe that significant machine interpretation of rough documents will require, at a foundational level, algorithms and representations capable of identifying and labeling visual structure comparable to human perceptual processes.

1.2 Tasks

The range of tasks that people perform around documents is not limited to straight reading. We skip around and visually navigate the document, we high-



a



b

Figure 1: Rough and formal versions of mixed text and graphics content.

light and annotate, we copy and edit selected material, we file and organize based on the document's appearance, style, format, and content. Computer systems for managing rough documents should support such a range of tasks. Even when classical recognition cannot succeed, for example in transcribing sloppy handwriting or an unfamiliar language, image analysis tools can still be useful for other purposes. Examples include:

- *Sorting* and classifying documents according to genre or other properties. For example, without attempting to actually transcribe the circuit, the visual qualities of Figure 2d indicate that it is a schematic diagram, it is hand-drawn, it is drawn on notebook paper, and it contains some titular text. By contrast, again prior to performing any text recognition, 2g has the visual characteristics of a receipt, with logos and a watermark.

- *Excerpting* significant sections of rough documents, for example separating the scattered handwritten text of Figure 2e from the photographs, to be entered into a handwriting database. In Figure 2d, perceptual-level recognition of the parallel line structure is critical to the lifting of the writing and drawing from the background rule lines of the paper.

- *Annotating* and *cross-linking* in an electronic document management system would benefit from the ability to attach annotations and links to particular perceptually salient image items, which in turn should be available as instantiated image objects to the underlying system. Even more beneficial would be the ability to detect and flag potentially interesting visual events occurring on rough documents on the basis of their perceptual qualities, such as logos, stains, watermarks, stamps, fingerprints, signatures, and sketches.

- *Editing* image material in rough documents offers the ability to rapidly assemble new documents by re-using complex text and graphics found on existing documents, without having to convert to structured formats.

- *Transcribing* handwritten text and diagrams to ascii and structured graphics formats requires the ability to follow the flow of the text regardless of slant and uneven flow, and to recognize

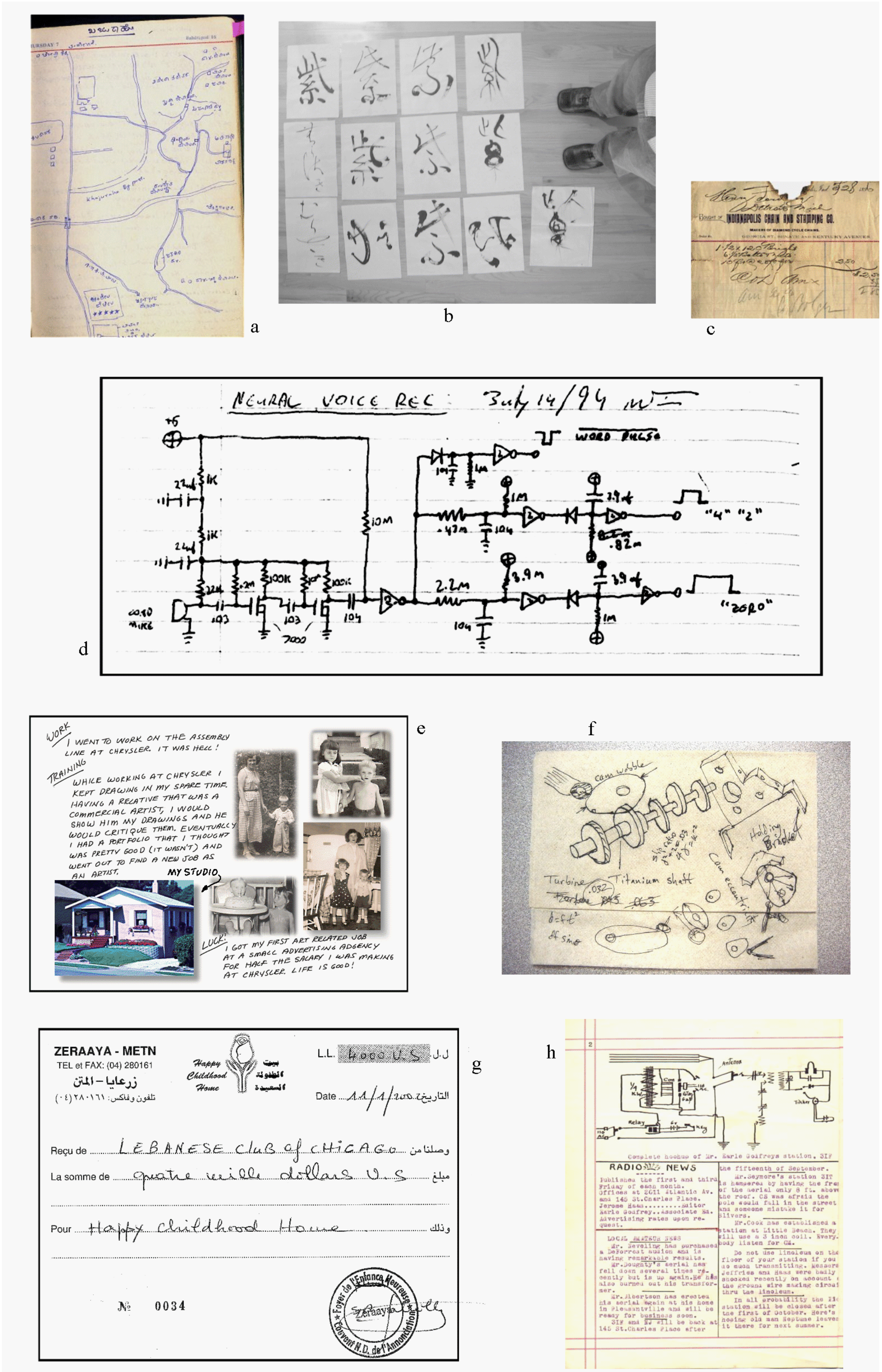


Figure 2: Rough documents arise from unpredictable sources and uncontrolled imaging conditions.

graphical structures despite their failure to obey strict graphical drafting rules.

- *Indexing and Retrieval* of rough documents would benefit greatly from automated means for labeling image content on the basis of visual qualities, even short of formal recognition in terms of natural language or graphical language constructs.

We are confident that exposure to practical application scenarios for document image management will lead to articulation and refinement of these non-traditional tasks supported by perceptual level image analysis of rough documents, and the invention of new ones.

2 Design Principles

2.1 Strong and Weak Models

Conceptualizations of both natural and artificial visual systems tend to be organized around a sequence of processing stages. The kind of circuits necessary at the sensory levels to manage dynamic range and contrast variation are different from the algorithms employed for, say, feature detection, which in turn are different from the computations involved in classification, model matching, and search.

Technology for recognition of document images bifurcates at an early stage between processing steps employed for text recognition and those employed for graphics recognition, respectively. These may have in common an image processing stage that may include thresholding, followed by a stage performing text/graphics separation. Then, text recognition processes typically follow the stages of skew detection, page layout analysis, word and character segmentation, and character symbol classification. Wherever possible, lexical or other semantic constraints are used to resolve ambiguous classifications and improve accuracy. Graphics recognition processes typically follow the stages of graphics vectorization, symbol matching, and knowledge-driven parsing of the vectors in terms of domain models particular to the drawing domain [Tombre and Chhabra, 98]. Examples of domain models include drafting rules for drawing mechanical parts and their dimensioning information, the syntactic rules governing the components and wires of electrical schematics, and architectural conventions for arranging walls, doors, windows, and furniture.

Both text and graphics lines of processing rely heavily on *strong prior models* of document image content to resolve ambiguities, to constrain possible interpretations, and to delimit and guide successive processing steps. These assumptions are appropriate when applied to documents whose image content is in fact confined, for example to pages of printed text or scans of engineering drawings. But the larger document image domain that includes rough documents cannot be approached in this way because the image quality, stylistic aspects, layout, formatting, and content of rough documents are so variable and unpredictable. Human perception is capable of detecting and making use of visual structure that violates standard norms defined for proper text layout and graphical drawing rules. Machine vision systems for rough documents must have this flexibility as well.

2.2 Knowledge Levels

Such a perspective leads our research group to a five-level framework for document image analysis that modularizes knowledge in a different way. This is illustrated in Figure 3.

1. Early Processing: Early visual processing encompasses image processing and initial stages of feature detection. Electronic images of documents captured not from scanners but from digital camera images of notebooks, whiteboards, flipcharts, projected images, etc. will in general carry artifacts of the imaging process due to sensor noise, sampling, focus issues, geometric distortion, dirt, smudges, illumination color and intensity gradients. These cause image degradations that impede extraction of the truly significant markings. One function of early processing is to "clean up" or condition the input using various image processing techniques such as color normalization, contrast equalization, adaptive thresholding, despeckling, etc. A second function of early processing is to identify primitive visual features such as lightness intensity edges, ridges, and solid regions.

2. Perceptual Organization: The Gestalt psychologists of the early 20th century identified a number of principles that characterize spatial structure the human visual system is adept at detecting. [Wertheimer, 23, Kanizsa 79, Green, 00]. Among the most important of these "Gestalt laws" are curvilinear alignment (or smoothness), closure, spatial proximity, feature similarity, and symmetry. These properties are viewed as being important clues to object cohesiveness and identity in natural scenes, and they are found to strongly govern the perception of synthetic images as well.

Because documents as created by people are intended to be seen and interpreted by our visual systems, it is likely that their structure naturally tends to reflect the fundamental visual capacities that we are endowed with. Indeed, Figure 4 illustrates Gestalt laws by way of positive and negative examples that respectively validate and violate our

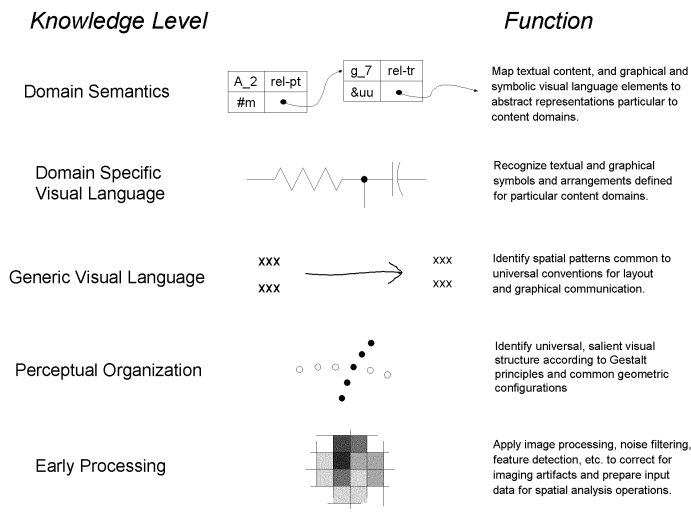


Figure 3: A Research framework of five levels of analysis for rough documents.

automatic perceptions. One rarely encounters a document whose semantic intent is depicted spatially in violation of the principles of perceptual organization. And when we do, these documents are confusing and difficult to read.

We believe that visual Perceptual Organization (PO) has an important role to play in computational models for rough documents. Starting with an undifferented array of pixels, the purpose of this stage is to identify chunks, groupings, and patterns in image material reflecting visually salient structures that are likely to reflect syntactically and semantically meaningful elements and relationships. Short of committing to conclusive mappings to the document's semantic domain, PO provides a rich set of candidates and building blocks to draw upon in constructing these mappings. Moreover, the knowledge required to do this processing, i.e. the data structures and algorithms required to in essence "implement" the Gestalt laws, would apply universally, across all document content domains, spanning both text and graphics.

The principles of PO are supported by many extremely compelling psychophysical demonstrations, but have proven very difficult to formulate as generally applicable computational algorithms. This remains a subject of active research in Computational Vision [Witkin and Tennenbaum, 83; Sarkar and Boyer, 94; Boyer and Sarkar, 99; Boyer and Sarkar, 00; Jacobs and Lindenbaum, 01].

3. Generic Visual Language: In addition to implicitly respecting the principles of human perceptual organization as it supports visual interpretation of natural scenes, we believe that human graphic communication embodies another level of structure consisting of standards and conventions that are culturally evolved and culturally acquired. In our hypothesis, this *Generic Visual Language* (GVL), serves certain universally important functions in communication such as demarking distinct groups, indicating progressions or other relationships, labeling, referring, and so forth. Some conventions of GVL involve graphical elements or symbols, such as lines, arrows, and encirclings, while others pertain to spatial arrangements, such as paragraph structure, titles and captioning, tabular structure, hierarchical indentations, and the proximity of text labels to graphics objects they refer to. Examples are shown in Figure 5.

The notion of Generic Visual Language, and a detailed account of its constituent patterns and rules, has not been a focus of study in the literature, although a few researchers have examined aspects of this topic [Bowman, 68; Bertin, 83; Tversky, 95; Tversky, 00; Tversky et al, 00]. We anticipate that the challenges of characterizing the range, diversity, and universality of Generic Visual Language will prove a rich area of investigation. As with Perceptual Organization, this research must span a range of disciplines and activities, including gathering and cataloging of data, experimental studies with diagram users, development and implementation of algorithms for recognizing GVL constructs in representations of images, building cognitive models for how spatial and symbolic elements contribute to reading of diagrams, etc.

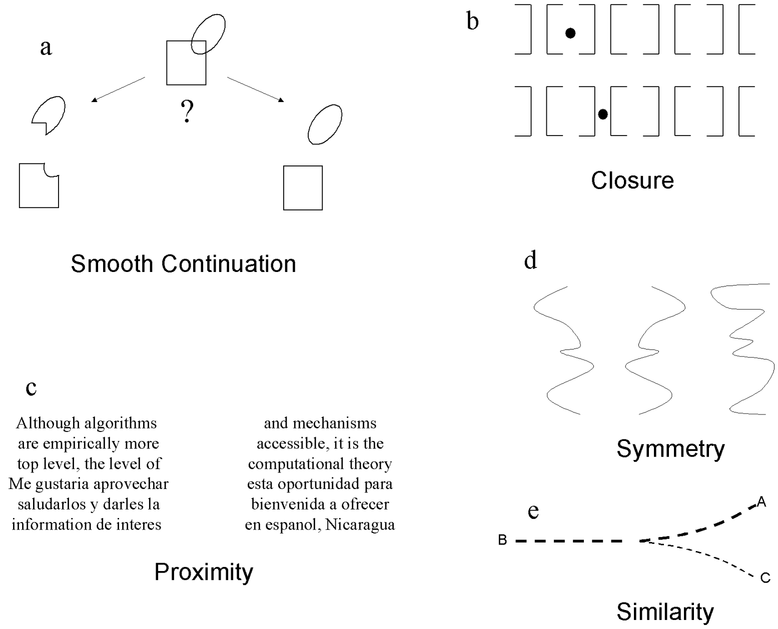


Figure 4: Illustrations of five Gestalt principles of visual perceptual organization. a. The central figure appears to be a combination of parts having smooth boundaries. b. The top dot appears to fall on a foreground object defined by a nearly closed boundary contour, while the bottom dot appears to lie on the background. c. The apparent visual partitioning of the text based on proximity overrides a partitioning based on semantic content. d. The curves that appear to go together are the pair forming a bilateral symmetry. e. The curve appears to continue from point B on the basis of similarity of its local properties.

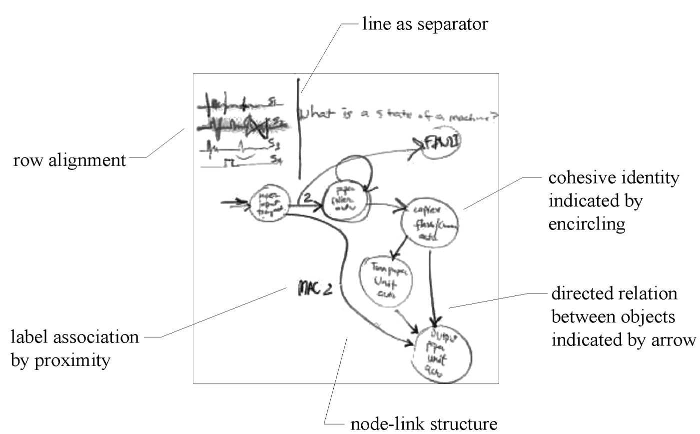


Figure 5: Examples of Generic Visual Language: visual patterns and notational conventions that convey formal or informal relationships, found across content domains.

4. Domain-Specific Visual Language: Whether or not universal conventions for spatial communication play a significant role in interpreting documents, it is clear that particular domains of discourse have developed their own specialized vocabularies and symbologies. Consisting of specialized symbols, linework, layout templates, and syntaxes for arranging these, *domain-specific visual languages* (DSVLs) are numerous and varied. We assume that all Domain-Specific Visual Languages respect the principles of Perceptual Organization and that PO-based segmentation and grouping will contribute to computational models for interpreting documents in terms of these DSVLs. Generic Visual Language conventions may provide a foundation and a resource for specialization in specific domains. For example, labeled arrows (lines with arrowhead symbols) may indicate a directed relationship between "from" and "to" object nodes, but particular domains may enforce rules such as the number of arrows allowed to enter or exit nodes, whether arrows may cross one another, and so forth.

5. Domain Semantics: While documents themselves are by nature spatially constructed representations renderable as images, the concepts and relationships they refer to may not be. Therefore, we assume that document interpretation systems may ultimately interact with knowledge and procedures at a non-spatial level, which we may call *domain semantics*. Some of this apparatus may be devoted to mapping between representations at this level and the domain-specific visual language level of description, while other machinery may be purely abstract or propositional and have nothing to do with document images at all.

Within this five-stage hierarchy, the Perceptual Organization and Generic Visual Language stages emphasize the foundational significance of *weak prior models* for document image analysis, that is, to identify meaningful visual structure without resorting to strong assumptions about what will be found. The proposed emphasis is on building tools that will deliver at least *some* level of meaningful interpretation across a wide range of document image input, as opposed to performing very well a narrow class of images but failing on any input that falls outside the class. Such a foundation is likely to support not only the interpretation of rough documents, but formal documents as well, especially when they contain rough aspects, such as handwritten annotations, logos, tables, and figures.

3 Technology and Research

Our group has engaged a number of topics focusing on the three earliest knowledge levels in this framework. Our efforts include:

- Image Processing
  - mosaicing of multiple snapshots for a video camera-based whiteboard scanner
  - page image dewarping for a face-up book scanner
  - color normalization for documents imaged with a camera
- Perceptual Organization
  - curve saliency and grouping
  - sketch recognition
  - occluding surface labeling
  - closed path detection
- Visual Language Recognition
  - structural page description and matching
  - structural recognition by constraint-based subgraph matching
  - sketch recognition exploiting PO-based data graph rectification
- Applications
  - perceptually-supported digital ink sketch editing
  - diagrammatic user interfaces
  - physical/virtual collaboration surfaces
  - perceptually-supported bitmap-based document image editing

In the following sections we touch on a few of these projects.

3.1 Image Processing

The advent of digital cameras increasingly enables casual document image capture. The resulting images suffer from artifacts not encountered through traditional scanning. One important image processing function is to correct for uneven illumination of the document. Our algorithm for lightness correction is based on inverting the image equation relating illumination, I, surface reflectance R, observed lightness L,

$$L = IR$$

If one is able to construct an estimate of the illumination I, then the underlying document reflectance can be recovered. A standard approach attempts to recover illumination under the assumption that most of the document consists of background white, on which dark markings are sparsely distributed. Under this model, foreground markings can be detected by high-pass filtering. The lightness values of the background can be interpolated across the resulting foreground mask.

However, this approach fails when the document contains large regions of foreground markings, such as contained in business graphics and artwork. Figure 6b illustrates the results of the commercial program, Whiteboard Photo. The fundamental problem is that, because illumination may vary across the image, it is impossible to determine locally whether the RGB value of any given patch is due to colored illumination off a white background surface or reflectance of a bright illuminant off a marked foreground surface.

To address this problem we have recently developed enhanced methods for detecting larger foreground regions, leading to improved color-normalization as exemplified in Figure 6c. We believe that much research remains to be done toward global spatial integration of image cues about lightness and reflectance changes which will lead to a comprehensive solution to this problem.

3.2 Perceptual Organization

The field of computer vision has seen some progress in formalizing middle-level visual processes of Perceptual Organization as computer vision algorithms for such processes as texture segmentation[Malik et al, 99], curvilinear line aggregation[Amir and Lindenbaum, 98], shape decomposition[Johannes et al, 01], and closed path detection[Jacobs, 96], among others. Because of limited efficacy, suitability of image content, and computational cost, rather little of this research has been carried over directly to document image analysis. However, the overall motivation and certain algorithmic techniques for finding perceptual structure in document images has taken root in the document image analysis community. Examples include grouping approaches to page segmentation [Kise, 98], grouping of curvilinear strokes in maps [Thomson and Brooks, 02], and grouping of text elements in engineering drawings [Kasturi et al, 92, Burge, 03].

One core paradigm is a three-stage process consisting of the following steps:

1 *Segment* the image into primitive elements.

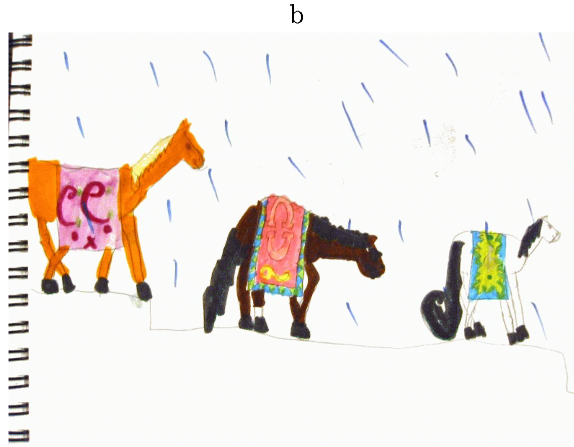
2 Form *link* relations among related elements.

3 Perform *search* for collections of linked elements that satisfy target criteria.

Different kinds of visual structure are detected according to specific criteria, parameters, and algorithms employed for each of the steps.

a

b



c

Figure 6: a. Original digital camera image. b. Processing by the commercial product, Whiteboard Photo. Note degradation of the solid color regions. c. Result of our algorithm which detects large foreground regions.

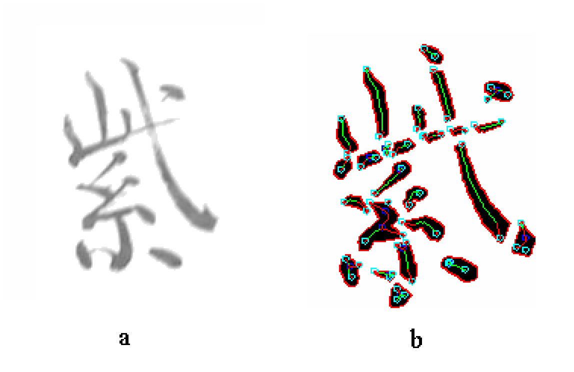


Figure 7: a. One of the ideographs extracted from Figure 2b. b. Exploded view of elements resulting from splitting to extract salient curvilinear fragments.

3.2.1 Segmentation

In document image analysis, the segmentation step almost always starts with the extraction of connected components in a binarized image. In clean images of formal documents, individual characters are frequently segmented into distinct components. Line drawings, on the other hand, undergo a further step of vectorization, which delivers units corresponding to straight or curved segments. Very little work has been devoted to articulated segmentation of solid graphic regions, although some commercial systems include detection of photographic regions which are extracted as distinct image objects.

Traditional analysis of degraded documents addresses the problem of distinct semantic image objects touching one another and thereby being inseparable on the basis of connected component segmentation. For text, special algorithms have been developed to perform character segmentation, while some limited amount of work has been done on separating touching symbols and line-art in technical drawings [Shimotsuji, et al, 94].

The problem of separating touching objects is especially important when addressing rough documents, which frequently include hand-drawn sketches and handwritten notes. Our group has focused some attention on this problem, and our approach has evolved over time, starting from one based on iterative classification of connected components and splitting off of curvilinear fragments of limited thickness [Saund, 02]. More recently, we have developed techniques that combine information derived from contour features with scale-invariant detection of "strokes" in order to split off significant curvilinear objects at perceptually natural cut points. Figure 7 presents an example.

3.2.2 Linking

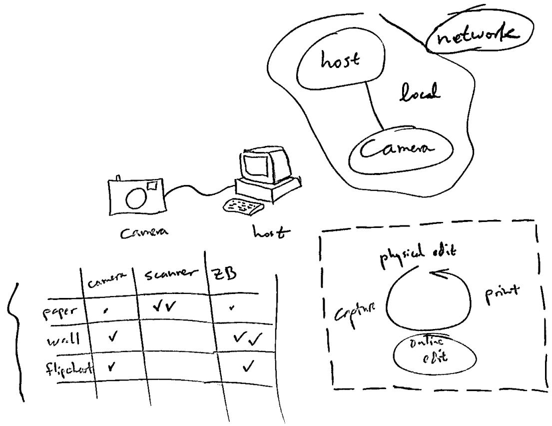
The purpose of linking is to form a graphical substrate of candidates for grouping collections of elements. Linking of image objects is often done based on spatial proximity (e.g. k-neighbors) or the Delaunay graph. In their simplest form, links are purely categorical, all-or-none. Greater use of links is gained by endowing them with attributes reflecting local spatial properties such as distance and orientation between respective pairs of linked objects.

Our research has pushed this idea further by iteratively refining the set of links according to the local spatial environment. In [Mahoney and Fromherz, 02] for example, the initial sets of linked curve ends are augmented to include the transitive closure of any set of links. In [Saund, 03], linking criteria are adjusted adaptively to deliver sets of links of bounded degree. Both this work, and [Saund, 99], each introduce links with rich type attributes which are in turn attributed by local measures of spatial configuration.

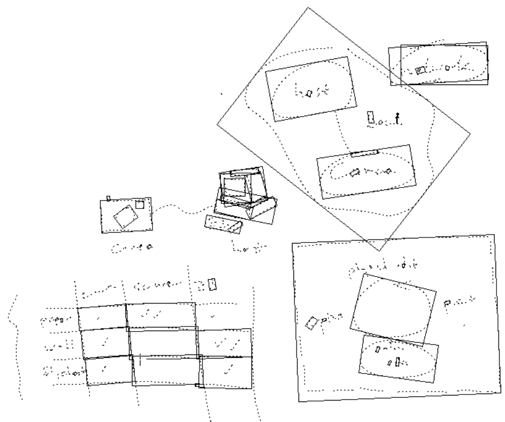
3.2.3 Grouping Search

Under the segmentation/linking/grouping framework, grouping amounts to selection of subgraphs of the link graph. The simplest version of this is the selection of transitive closures, assuming initial links have been pruned down to exclude links between elements that should not be grouped together, leaving "islands" of connected components in the graph. This is equivalent to tracing in one dimension, and coloring in two dimensions. More generally, graph partitioning algorithms [Shi and Malik, 00] or search algorithms are applied to generate sets of linked elements that reflect target global criteria. Thus far, these criteria mainly reflect spatial proximity, curvilinear alignment, and path closure. Our group has developed versions of all of these grouping criteria. For example, Figure 8 shows a result from our recent work on the identification of perceptually closed paths in line drawings [Saund, 03]. This technique involves two primary innovations. First, we identified criteria relating both the local progression of lines through junctions, and the global shape of paths, to paths' perceptual salience. Second, we showed that a bi-directional search procedure ameliorates garden path searches and permits qualified paths in the link graph to be found efficiently.

In general we believe that the issues of forming and refining link graphs, and formulating search criteria and search algorithms to identify perceptually salient visual structure, will continue to be key areas of research important to sensible machine interpretation of rough documents.



a



b

Figure 8: a. Sketched figure. b. Closed paths found by our bidirectional search algorithm.

3.3 Visual Language Recognition

Perceptually salient collections of primitives labeled as groups at the PO stage represent resources, or building blocks, with respect to which recognition procedures may perform matching to structural models. Under this framework, much latitude remains in the content of a structural model database, and in the matching procedures used.

Under the five knowledge levels identified in Section 2.2, we anticipate that many if not most rough documents will reflect general graphical conventions that may eventually be expressed as a relatively stable database of generic visual language constructs. Elements of generic visual language include the various roles that straight lines, arcs, encirclings, arrows, and proximal text play as communicative devices for grouping, separating, labeling, referring, indicating progression, indicating logical structure, indicating text flow, and so forth.

By contrast, model databases for domain specific visual languages pertain to specific document domains such as journal layouts, schematic diagrams, tables, charts, schedules, technical drawings, and mathematical notations. These databases may be expected to grow and adapt to different application domains, which may nonetheless span both rough and formal expressions of underlying semantic content.

Our research adopts the strategy of seeking a common structural matching framework which will apply to both Generic and Domain Specific visual language.

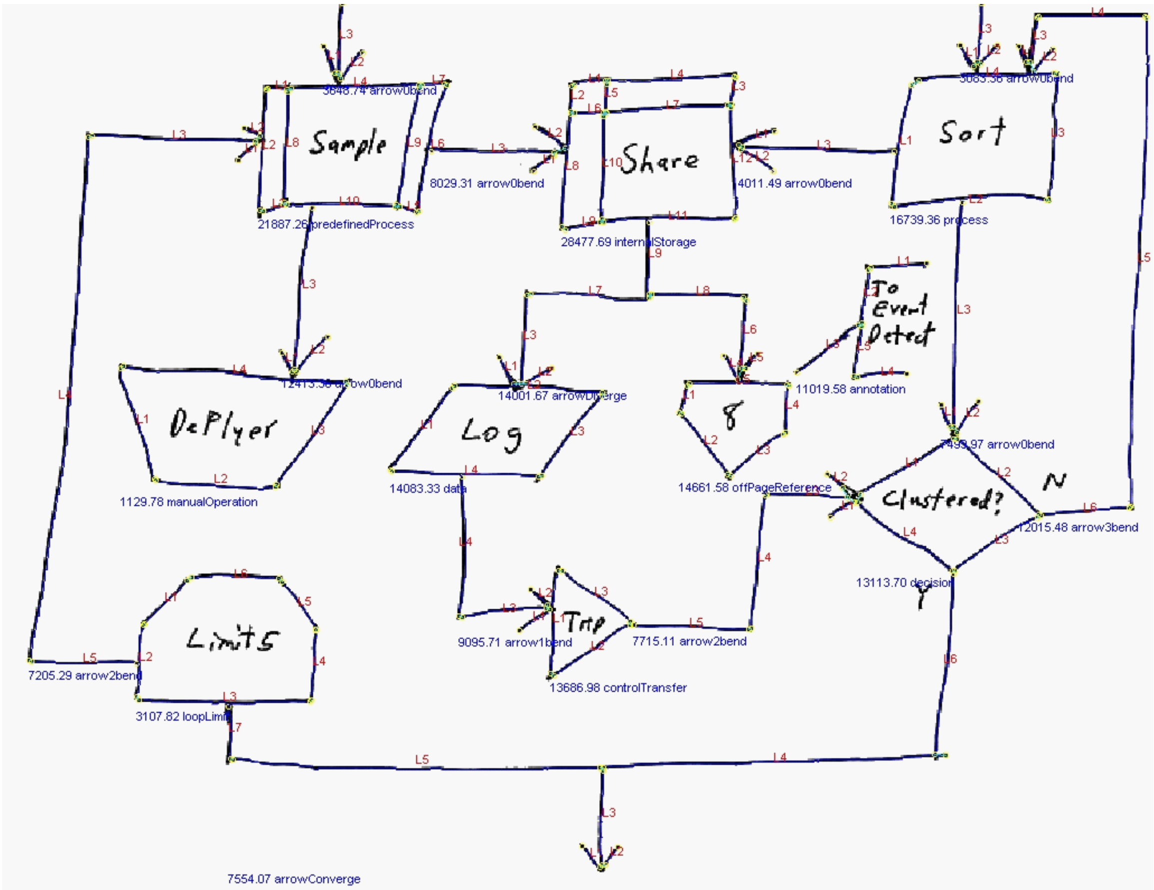
3.3.1 Structural Modeling and Matching

Structural matching recognizes and parses graphical configurations. By "parsing" we mean individually associating each input mark with the model part that it depicts. Because of ambiguity in the association between model objects and data objects, this process must cope with the multiplicity of alternative local groupings produced by the Perceptual Organization stage, and it must be capable of producing alternative plausible interpretations of the same data, ranked by their plausibility. It must cope with the inherent exponential worst-case complexity of structural matching in the graphical domain, and it must be easily extensible with new configuration models.

Our group is developing a subgraph matching formulation, implemented in a constrained optimization framework. See Figure 9. The PO stage generates multiple plausible groupings leading to alternative hypotheses for assignment of model parts to observed data. One set of constraints enforces a unique selection among competing alternatives. The modeling language for specifying configurations allows topological and geometric constraints, both hard and soft, on the spatial relationships among parts that define model objects. These constraints feed in a straightforward way into an overall measure of match quality that serves as the objective function for optimization. We believe that this approach will elegantly handle issues of ambiguity, and provide a solid foundation for dealing with the issues of complexity and open-endedness.

4 Application Platform: Perceptually Supported Image Editing

We are pursuing these lines of research in Perceptual Organization and Visual Language Recognition in rough documents against an application platform that simultaneously tests, and delivers value from, our research as it progresses. This platform, implemented in Java, presents a toolkit for building graph-



a

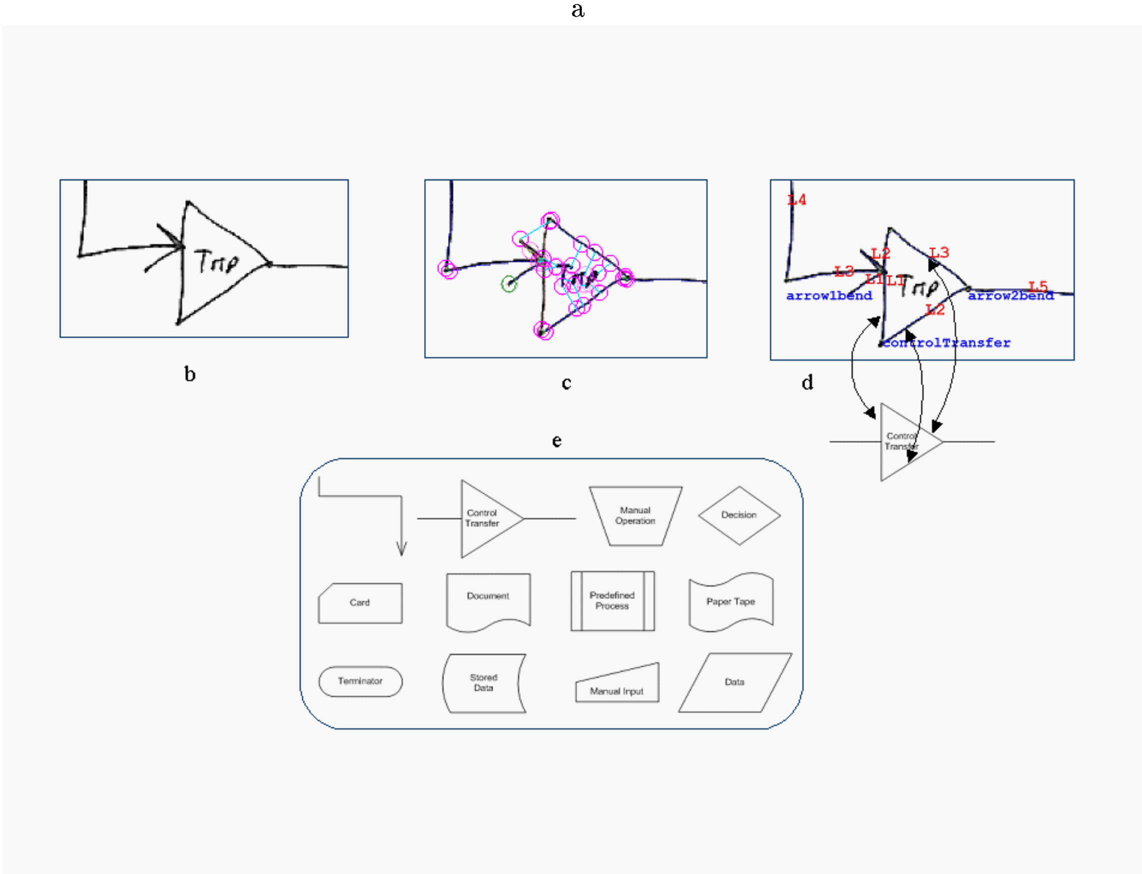


Figure 9: a. Hand-drawn diagram tagged with labels indicating matched parts. b. One item from this diagram. c. Spatial relations among features. d. Corresponding parts found by matching to one item in the model database, e.

ical user interfaces to applications calling upon our research code base.

One application, called *ScanScribe*, is a perceptually supported document image editor. At its basic level, ScanScribe offers selected functionality found in photographic image editors like Photoshop, but with an interaction model targeted especially to easy selection and manipulation of image material in documents. In particular, color normalization processing described in Section 3.1 performs automatic foreground/background layer separation, in which the background is rendered transparent. Then, as the user selects and manipulates image objects present in a single input bitmap, they retain their identity as independent bitmap fragments which can be dragged, duplicated, grouped with one another, and carved into yet smaller pieces. By design, this process occurs automatically and below the user's conscious awareness.

While it is quite novel and useful at its base level, ScanScribe presents a platform for testing and validating image analysis research. The ScanScribe architecture is designed to make available the results of any grouping processes that automatically recognize visual structure—at whatever level—in the image. For example, if a recognizer were available that recognized tabular structure, ScanScribe could present the user with easy ways to access, at will, the rows, columns, individual elements, and entire table. Because a user-interactive image editor puts the user intimately in the loop, even partial or imperfect recognition processes can become useful. This is especially significant when working with rough documents, many exemplars of which will prove extremely difficult for machines to interpret without errors for quite some time to come.

Because image editing is a broadly useful functionality, we anticipate that the ScanScribe document image editor may serve as a springboard or base level upon which other, more specialized applications, will be built. For example, a scenario in which a user wishes to perform database search on the ideographs in Figure 2b, requires the selection of this target from the other material in the image which could misdirect any database search processes.

5 Conclusion

Success in document image analysis has to date largely relied on having strong prior models of the input data, including strong models for scanner degradation processes, page layout, font, text content, engineering drawing rules, and so forth. These strong models are inappropriate for rough documents such as handwritten sketches and notes, graphical figures, annotated documents, complex layouts, and casually imaged documents. Instead, we advocate the development of core visual analysis competency at the levels of Perceptual Organization, dovetailing with Image Processing below and Visual Language Recognition above. The visual structure brought forth at these stages will support not only transcription, but will open a broad range of document image processing, analysis, interpretation, and management tasks.

6 References

Amir, A., and Lindenbaum, M., "A Generic Grouping Algorithm and its Quantitative Analysis," *IEEE TPAMI*, 20:2. 1998.

Bertin, J. *Semiology of Graphics: Diagrams, Networks, Maps.* Univ. of Wisconsin Press, 1983.

Bowman, W.J., *Graphic Communication.* J. Wiley, New York. 1968.

Boyer, K., and Sarkar, S., eds. *Perceptual Organization for Artificial Vision Systems.* Kluwer. 2000.'

Burge. M., "Document Image Analysis: Vectorization, Segmentation, Classification & Grouping," "http://www.computing.armstrong.edu/FanNStaff/burge/java/grouping/grouping.html"

Green, C., "Introduction to: 'Perception: An introduction to the Gestalt-theorie' by Kurt Kaffka (1922)", "http://psychclassics.yorku.ca/Koffka/Perception/intro.htm". 2000.

Kanizsa, G., *Organization in Vision: Essays on Gestalt Perception.* Praeger, New York.1979.

Jacobs, D., "Robust and Efficient Detection of Salient Convex Groups", *IEEE TPAMI*, 18:1. 1996.

Jacobs, D., and Lindenbaum, M., *POCV 2001: The Third Workshop on Perceptual Organization in Computer Vision*, CIS Report CIS-2001-05, Technion, Haifa, Israel. 2001.

Johannes, M., Sebastian, T, Tek, H., and Kimia, B., "Perceptual Organization as Object Recognition Divided by Two," in Jacobs, D., and Lindenbaum, M., *POCV 2001: The Third Workshop on Perceptual Organization in Computer Vision*, CIS Report CIS-2001-05, Technion, Haifa, Israel. 2001.

Kasturi, R, Raman, R., Chennubhotla, C., and O'Gorman, L., "An Overview of Techniques for Graphics Recognition", in Baird, H.S., Bunke, H., and Yamamoto, K., eds., *Structured Document Image Analysis*, Springer-Verlag, Berlin. 1992.

Kise, K., Sato, A., and Iwata, M., "Segmentation of Page Images Using the Area Voronoi Diagram," *CVIU* 70:3. 1998.

Mahoney, J., and Fromherz, M., "Interpreting Sloppy Stick Figures By Graph Rectification," in D. Blostein and Y-B. Kwan, eds., *Graphics Recognition: Algorithms and Applications*, 4th International Workshop, GREC 2001, Springer LNCS 2390., 2002.

Malik, J., Belongie, S., Leung, T., and Shi, J. "Contour and Texture Analysis for Image Segmentation," in K. Boyer and S. Sarkar, *Perceptual Organization for Artificial Vision Systems*, Kluwer. 2000.

Sarkar, S., and Boyer, K., "A Computational Structure for Preattentive Perceptual Organization: Graphical Enumeration and Voting Methods". *SMC(24)*. 1994.

Saund, E., "Perceptual Organization of Occluding Contours of Opaque Surfaces", *CVIU* 76:1. 1999.

Saund, E., "Method and apparatus for separating document image object types", U.S.P. 6,411,733. 2002.

Saund, E., "Finding Perceptually Closed Paths in Sketches and Drawings," *IEEE TPAMI*, in press.

Shi, J., and Malik, J., "Normalized Cuts and Image Segmentation," *IEEE TPAMI*, 22:8. 2000

Shimotsuji, S., Hori, O., and Asano, M., "Robust Drawing Recognition Based on Model-Guided Segmentation", in A. L. Spitz and A. Dengel, ets, *Proc. IAPR Workshop on Document Analysis Systems*, 1994.

Thomson, R., and Brooks, R., "Exploiting Perceptual Grouping for Map Analysis, Understanding and Generalization: The Case of Road and River Networks" in D. Blostein and Y-B. Kwan, eds., *Graphics Recognition: Algorithms and Applications*, 4th International Workshop, GREC 2001, Springer LNCS 2390., 2002.

Tombre, K. and Chhabra, A.K., (Eds). *Graphics Recognition - Algorithms and Systems.* Lecture Notes in Computer Science, vol. 1389. Springer Verlag. 1998.

Tversky, B., Zacks, J., Lee, P., and Heiser, J., "Lines, Blobs, Crosses, and Arrows: Diagrammatic Communication with Schematic Figures". In M. Anderson, P. Cheng, and V. Haarslev, eds. *Theory and Application of Diagrams*, LNCS 1998, Springer. 2000.

Tversky, B.: "Some Ways that Maps and Diagrams Communicate". *Spatial Cognition.* 2000.

Wertheimer, M.. "Laws of Organization in Perceptual Forms", In Ellis, W., ed, *A source book of Gestalt psychology*, Routledge & Kegan Paul, London. 1938 (1923).

A.P. Witkin and J.M. Tenenbaum. On the role of structure in vision. In J. Beck, B. Hope, and A. Rosenfeld. (eds.), *Human and Machine Vision.* Academic Press. 1983.