Document Image Retrieval Based on Layout Structural Similarity

Christian Shin[+], David Doermann[++]
[+]Department of Computer Science, State University of New York, Geneseo, NY 14454
[++]Language and Media Processing Laboratory, University of Maryland, College Pak, MD 20126
Email: shin@geneseo.edu

***Abstract*—In this paper, we describe issues related to the measurement of structural similarity between document images. We define structural similarity, and discuss the benefits of using it as a complement to content similarity for querying document image databases. We present an approach to computing a geometrically invariant structural similarity, and use this measure to search document image databases. Our approach supports both full image matching using query by example (QBE) and sub-image matching using query by sketch (QBS). The similarity measure considers spatial and layout structure, and is computed by aggregating content area overlap measures with respect to their underlying column structures. These techniques are tested within the Intelligent Document Image Retrieval (IDIR) System, and results demonstrating effectiveness and efficiency of structure queries with respect to human relevance judgments are presented.**

***Index Terms*—Document image understanding, indexing and retrieval of document images, similarity, document layout structure**

I. Introduction

SCANNED image archives of hardcopy documents are increasingly used either to replace paper and microfilm filing in an attempt to move toward a paperless office or to enhance the effectiveness of search. Typically these archives are combined with a database management system. Making full use of the capabilities of database indexing and retrieval techniques requires that images be processed and analyzed to adequately describe them, and that a mechanism be provided through which images of interest can be retrieved and presented efficiently and effectively. The effectiveness of any image database system depends on a number of factors including the image analysis techniques, the image content description, the storage of images and their attributes, the similarity or dissimilarity measures, the range of image queries allowed, the effectiveness of the search interface, and the efficiency of implementation [8].

In early image databases, images were often manually processed and analyzed to take advantage of automatic database organization, storage, and retrieval capabilities due to maturity of database technologies compared to image understanding technologies. Manual processing typically involved associating a set of keyword descriptors with each image. For large databases, manual indexing can be prohibitively expensive, not to mention the subjective and possibly myopic interpretation by the person creating the index, and the limited expressiveness of keywords. As a result, the problem of automated processing and retrieval of images by content has evolved as an active area of research.

For *scene* image databases, in order to support content queries images are analyzed to extract quantitative and qualitative descriptions of their content to be stored as index information. Current content-based image database applications formulate queries in terms of meaningful quantitative image features, such as color, shape, and texture, or provide an example image which resembles the desired result with respect to these features. These descriptions are application domain dependent and are used in the searches to generate a set of candidate images which may satisfy the given query.

The indexing of *document* image databases differs from the indexing of scene image databases in a number of ways. First, the lack of variation in basic image properties compared with, for example, color images of natural scenes, limits the use of previously explored techniques. One approach which has been used to define similarity between images is to look at the color histogram, and/or the locations of various color regions within the image [7]. Since most documents are black-and-white and in fact pseudo-binary, the use of color indexing techniques is for the most part meaningless. Second, the meaningful information in a document's layout is not necessarily contained in the shapes of local objects, but in the more global structure of the page. Therefore, while simple texture measures work well for detecting edge tendencies, for example, the fact that text typically falls horizontally in most documents makes such local measures less than ideal in document retrieval applications. Third, much of the information of interest is contained in the actual content (text or graphics) of the document, not just in the type of object present. For scene image databases, the existence of a particular object such as a house or tree may be a key index feature, but typically in document applications, the existence of a significant structural feature is qualified by some information about the content - for example, “find a table containing US Census data”.

The primary approach to indexing document image databases has therefore been to use OCR to get a content representation and rely on indexing the text directly. Retrieval of text documents that are similar with respect to content has been addressed by researchers in information retrieval for many years. Techniques however are highly dependent on the quality of the OCR. For certain domains conversion is a viable option, but in general, complete conversion is not possible for a number of reasons. First, OCR requires extensive computation time for image processing and classification, and may have a significant degree of recognition error on poor quality documents; this makes OCR an insufficient means for complete capturing of information. The quality can be affected by a number of factors including the physical medium used to create the document, and the way the document was represented. Some documents are handwritten and OCR technology has not progressed to the point where unconstrained handwritten text can reliably be recognized. Second disadvantage of OCR techniques is that the layout or formatting of the document is typically not preserved. As recognized by [1][2][5], users searching for a particular document in a large document database tend to rely on clues about the form and structure of documents. Such clues, which could be obtained from either the original bitmap image or reduced scale images (i.e., thumbnails), tend to be lost in ASCII text renderings of images. The layout or formatting of a document is crucial information that can be used to identify similar documents in a large database. Third, there are no usable forms expressible for graphics images, logos, etc.

Current document image database techniques (e.g., text-based search, content-based image search, etc.), however, do not typically take full advantage of the document's geometric and logical structure. Document structure can provide a great deal of information in determining which documents are relevant to a given query. In many applications, structure-based matching enhances existing content-based matching capabilities, and provides an effective way to quickly reduce the set of candidate documents for similarity matching using layout structure knowledge (e.g., location and extent of components, spatial relationships among the components) and logical structure elements (e.g., memo's to, from, subject, date). Structure-based matching is performed without a priori knowledge of document type, and without OCR. Such structure-based matching techniques can be used with traditional text-based information retrieval techniques to take advantage of structural information by constraining text-based search results with layout conditions (e.g. finding documents with a text string appearing in upper half of a page).

The fundamental search operation for traditional databases is exact or range matching of text strings. The data in the database should be the same or within a given range of values, in some predefined sense, as the query. With complex image data, exact match queries are difficult to express or cannot be expressed precisely, and we see the emergence of database systems in which the search operation attempts to retrieve portions of the database ranked by similarity (or dissimilarity) assessment, where similarity is a measure that is defined and meaningful for every pair of images in the image space. The need for an informative similarity or distance measure between objects or between their respective representations is essential in handling the practical aspects of retrieval. Measuring similarity between two images is particularly important in large image databases where limited supplemental metadata is available. The ability to handle similarity (or approximate) search is particularly desirable when (1) a user does not know exactly what he/she is looking for, (2) a user knows exactly what he/she wants, but the query is difficult to express or cannot be expressed precisely, or (3) there does not exist an exact match, but rather there exist a set of acceptably *close* matches. The important question in practical applications is not whether two objects are identical, but rather how similar they are to each other with respect to other entries in the database [4]. Although the most important operation in image databases is similarity-based searching, the least well-defined of these factors is perhaps the similarity measure.

In this paper we explore the concept of structural similarity. A discussion of what it means for documents to be structurally similar is presented in Section II. In Section II.C, we present our approach to structural matching and measuring structural similarity between images, and we apply these similarity criteria, measures, and indexing mechanisms to the task of retrieval from document image databases by using area overlap structural similarity measures. In Section III, experimental test results are presented, and conclusions and future work are discussed in Section IV.

II. STRUCTURAL SIMILARITY MATCHING

A. Definitions

In this section, we define terms related to measuring similarities between two entities in general, and we apply these definitions to measuring similarities between two document images.

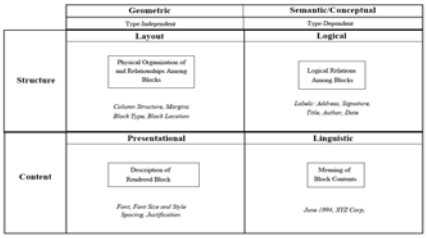

Figure 1. The geometric, and semantic content and structure descriptions.

An *object* is defined to be a spatially compact region of interest which is recognized as a single entity. The definition of an object can be made recursive by defining two types, basic and composite. A *basic* object is a lowest level (smallest) object with no structure of interest, and a *composite* object is an object that consists of one or more basic and/or composite objects, referred to as its

components. The organization of *component* objects is referred to as *structure*. The structure, for example, can be a grouping (e.g., a set or ordered list), or a hierarchical organization (e.g., a tree). The definition of an object is such that although it may be considered a basic object with respect to a given application, it can be viewed as a composite object in another application. The granularity of the basic object depends on the application, and it may vary over time, level of interest, or point of view (perspective). For example, for a reader of a document, a character may be a basic object, while someone studying handwriting may be interested in the stroke-level features.

Similarity is a measure of relatedness between two objects, and is a function of their structural similarities and content similarities. *Content similarity* is a measure of relatedness between the properties or attributes of the content of two basic objects, and *structural similarity* is a measure of relatedness between the organizations of two composite objects. The computation of content and structural similarity measures are typically dependent on the features describing them and the representation scheme used for their structures respectively.

B. Features

As described in Section I, keyword-based searches traditionally require manual indexing of the document image, and full text-based searches require prior OCR. Content-based searches using image features, such as color, shape, and texture perform a reasonable job in scene image databases, but they are not typically suitable for document image databases. These search methods are not effective ways of querying document image databases, nor are they necessarily natural ways to express the query. We need more natural ways of defining what makes images of documents similar. Since it is the user that has to be satisfied with the results of the query, it is natural to base similarity measure that we will use on the characteristics of human similarity assessment. In conjunction with the technological advancement in document image analysis and economic feasibility (e.g., fast processors, inexpensive storage, etc.) of creating large databases of document images, there is a tremendous need for robust ways to access the information these images contain by allowing abstract and conceptual queries.

There have been many attempts to study not only different aspects of the human perception of similarity in psychology [9][10], but also work practices in office environments in social anthropology [1]. By examining different work practices, Blomberg et al. found that a work practice can be supported with a system that is capable of searching and retrieving documents in a database by their types. For example, an attorney in a law office wants to find memos that contain the text string “fraud” appearing in the memo, or he/she wants to find letters that were sent by “John Doe” between January ‘04 and January ‘06. The referenced types in the above examples, *memo* and *letter*, are logical types, and their member component objects are logical objects. For example, a *memo* consists of memo sender, memo receiver, memo subject, memo date, memo body, memo copy, etc. as its logical component objects, and a *letter* consists of letter sender, letter data, letter body, letter signature, etc. In order to satisfy these types of queries, we need to know the logical type of a document and its component logical objects. Documents which are created electronically and become part of an electronic document management system can typically be searched for by logical structure as well as by content provided they are from the same system (i.e., a format). We deal with bitmap images of scanned hardcopy documents, however, that have no structural components that are immediately perceivable by a computer. To be searchable, the structure of a document image needs to be analyzed to identify its logical structure. Identifying the logical type of a document generally requires type-dependent models, which describe its component objects and their organization, and we lack such information.

The logical component objects are laid out in a unique way forming a particular layout structure. Where some types of documents are general in the sense that they recur across different organization and work processes, other types of documents are specific to a particular user. For example, a business letter and a memo are examples of general classes. A set of documents with an individual's own design is an example of a type that is specific to a particular user, such as an advertisement flier or a poster. In addition, it has been found that many different types of documents have a predefined form or standard set of components that depict a unique spatial arrangement.

In order to maximize the transfer of information to the reader of a document by using vision as a medium, documents are designed in accordance with basic perceptual principles such as the principles of Gestalt [6]. For example, using white spaces as separators follows the principle of proximity, which states that elements which are closer together tend to be grouped together, thus forming zone (or block) boundaries. The principle of good continuation, according to which elements that lie along a common line or smooth curve are grouped together, causes the white spaces that border a column to be seen as units, thus separating the column from its neighbors. The principles of similarity, which states that elements that are similar in physical attributes, such as fonts, color, orientation, or size, are grouped together, causes zones with the same content type to group together forming content zones (or blocks) [2].

We need a method based on layout structures of document (i.e., visual appearance) to facilitate the search and retrieval of a document stored in a heterogeneous database of documents. Unlike many techniques for searching the text within a document, searching documents according to their layout structure is based on the appearance and not the textual content found in a document. The general premise for searching documents based on their layout structure is that the layout structure of a document often reflects its type. For example, business letters are in many ways more visually similar to one

another than they are to magazine articles. Thus, a user searching for a particular document while knowing the class of documents is able to more effectively narrow the group of documents being searched.

Authors typically use combinations of layout and content visual features (e.g. bold font for emphasis) to convey an intended organization, or to assign priorities to specific components. There is a level of document organization, which can be regarded as intermediate between the layout (geometric) and logical (semantic) levels, that relates to the efficiency with which the document transfers its information to the reader. This level is described as the functional level [2]. The functional description of a document is often independent of document type and can be derived from geometric/layout considerations. Headers, footers, lists, tables, and graphics are examples of generic structures which can be common to many types of documents. For example, memos are divided into memo header (having a list-like structure in the beginning of a page), memo body (having normal text blocks in the middle), and memo copy (having a small text block at the end of the page) as their functional components. In the absence of type-dependent models, we use functional-level organization of documents instead of type-dependent logical structure.

Often the layout of a particular document contains a significant amount of information that can be used to identify a document stored in a large database. In addition to the retrieval scenarios with logical types, the layout or spatial structure is useful on its own in such applications as a personal filing system where images are viewed and archived by an individual, and later retrieved by visual recollection. Examples include finding pages with a table on top and a graph on the bottom of a page, and finding three-column text with a graph in the middle of the page.

Structural similarity, in addition to the currently used content similarity, has an advantage of quickly narrowing down candidate images before expensive content matching is performed, and the structure can be used to help navigate content-based indexing. Structure-based matching is generally performed without a priori knowledge of a document type, and without OCR. This structure-based matching technique can be used with the traditional information retrieval techniques to take advantage of structure information by combining it with full text retrieval to search the document. For example, text searches can be constrained by geometric and functional features, such as "find documents with a title that contains the text string 'presidential election'," (a title is functionally defined as being located in top half of a page, and having greater than normal point size for the document, for example).

What are good structural similarity features? We find that relative spatial layout features (relative location, size, and aspect ratio) are more important than absolute geometry in searching for structurally similar images. The column structure plays a significant role, and the numbers and types of components present on a page as well as how they are distributed spatially are important. On the other hand, the use of geometric invariance in many image matching applications suggests that absolute scale, position (translation), and orientation are not as important in measuring structural similarity as are relative scale, location, and orientation.

C. Matching Algorithm

In this section, we present algorithms for computing the structural similarity based on our proposed approach. The structural similarity of two documents is measured by computing area overlaps of their constituent regions and their types (text, graphics or image). For each region R_i^Q in the query image Q, we match R_i^Q to each region R_j^D of the database image D of the same type that overlaps it. If there is no region of overlap, the region R_i^Q is mapped to NULL. We therefore have a directional graph from each query image region to a possibly empty subset of the database regions.

Once this first correspondence has been established, an evaluation mechanism is used to refine and measure the quality of the match. It is clearly possible for a single region in the query image to be mapped to multiple regions in the database image and vice versa. There are several situations where such a mapping is not desired, and must be refined.

The first restriction is that no region should be mapped to two or more regions in the horizontal direction. This would occur, for example, if a page with a single block of text were mapped to a page with two columns of text. Splitting a block horizontally may occur between paragraphs, for example, but vertical splitting is typically an intentional structural occurrence. For query regions which map to more than one corresponding database region, a subset of regions which have maximal intersection, but do not neighbor horizontally, is chosen and the remaining regions are removed from the mapping. For query image regions which overlap a single database region, the correspondence is trivial (but we must later consider a symmetric case where multiple query regions correspond to a single database region).

Figure 2

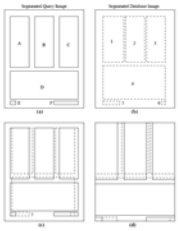

Figure 3

Figure 2. Structural similarity matching: (a) the segmented query , (b) the segmented database image. Figure 3. Structural similarity matching: (a) the segmented query image, (b) the segmented database image, (c) overlay of the two structures, and (d) normalized overlay of the two structures.

Once this condition is satisfied for all query regions, the symmetric case, where a single database region corresponds to multiple query regions, is still possible. Using the restricted mapping, a reverse mapping is constructed from the database image to the query image and the condition is

checked again. Regions which violate the vertical split are again evaluated and the subset of maximal overlap is kept.

Once the best match is found, the percentage of each region in the query image which matches is computed, and the total is summed for all regions in the query image. The same is done for the database image. For full image matching, the overall structural similarity is the minimum percentage of the overlap between the two sets of mappings, and for sub-image matching, the maximum is the overall similarity.

D. Matching Example

In this section, we walk through a complete example of the steps involved in computing the structural similarity between two document images (query and database images). Page segmentation is performed on these images, dividing each page image into a set of content zones; the segmented results are also shown in Figure 2. These segmented content zones have attributes including their geometric features, such as location and dimensions, in addition to their content properties. The geometries of the segmented zones from two pages are normalized to provide geometrically (scale and translation) invariant search capabilities. Figures 3 (a) and (b) show the zones recognized by the page segmentation process, (c) shows overlay of the two documents' structures, and (d) shows normalized overlay of the two structures.

Table 1. Mappings from query to database images.

Region	Mapping	Mapping with H-Restriction
A	A → 1	A → 1
B	B → 2	B → 2
C	C → 3	C → 3
D	D → 4	D → 4
E	E → 5	E → 5
F	F → 6	F → 6

Table 2. Mappings from database to query images.

Region	Mapping	Mapping w/ H-Restriction	Comments
1	1 → A	1 → A	
2	2 → B	2 → B	
3	3 → C	3 → C	
4	4 → A,B,C,D	4 → C, D	A & B removed
5	5 → E	5 → E	
6	6 → F	6 → F	

After preprocessing steps, we applied the algorithm for the structural similarity measure presented in Section II.C to two images in Figure 2. From Figure 3(d), we first generate two sets of mappings that show zones that have overlap areas between zones in the query image and database image structures, and vice versa. As described in Section II.C, when multiple blocks are overlapped with a single source block from different column structures, we select the set of blocks from a single column that produces the maximum area as a contribution to the overall area overlap structural similarity. These overlaps are represented as a set of mappings for each direction of the matching. The mappings are shown in Tables 1 and 2.

Area overlaps are computed for both sets of mappings. The structural similarity (a ratio of total area overlap to total area) for the first and second set of mappings are 0.82 and 0.95, respectively for the example above. Given the two structural similarity measures, the minimum value, 0.82, is the measure of structural similarity for full image queries, and the maximum value, 0.92, is for sub-image queries.

Figure 4. Query images.

III. RETRIEVAL EXPERIMENTS

In this section, we present an evaluation of the algorithm presented in Section II.C using our retrieval system, the IDIR, presented in [3]. In order to evaluate the performance of the algorithm, we used it to compute similarity scores between the 979 images in the UW-I image database and the twelve prototype images (shown in Figure 4). We then compared these similarity scores to the relevance scores derived from the human subjects' judgments for those same images, which we presented in [11]. Because of the difficulty of defining similarity even in a relatively restricted domain such as document images, in [11] we described the results of a study in which human subjects were asked to judge the similarity (here called "relevance") between a set of document images and a set of examples of document image classes.

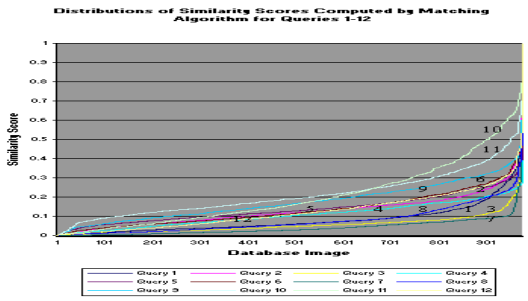

Figure 7: Similarity scores computed by the matching algorithm for queries 1-12.

We used the IDIR query interface, which we presented in [3], to perform retrieval by example by providing each of the 12 query images as seed images. For each of these queries, the entire UW-I database was ranked using the document page matching algorithm. The IDIR system produced 12 lists which resulted from matching each query image to the entire UW-I image database. The similarity scores computed by the algorithm for all 12 queries are plotted in Figure 7. The scores are sorted in ascending order from 0.0 to 1.0.

We then compared the scores obtained using the matching algorithm with the relevance scores obtained from the human subjects' relevance judgments. The relevance scores were computed by multiplying the number of judgments of relevance by the average of the relevance ratings assigned by all seven subjects, as described in [11]. Figure 8 shows the distributions of these relevance scores for each of the 12 query images for all 979 database

images. The scores are sorted in ascending order from 0 to 35. and the consensus relevance scores.

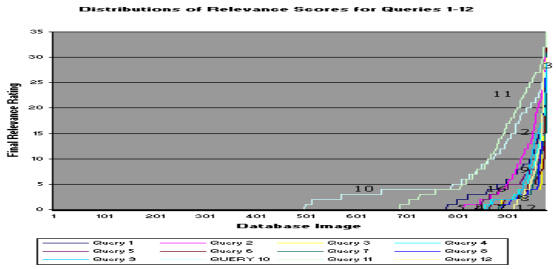

Figure 8: Relevance ratings given by the human subjects for queries 1-12.

Table 4: Correlation coefficients. The first column ("Alg.") shows the correlations between the similarity scores and the consensus relevance scores; The second and third columns show the standard deviation and the mean of the the following seven correlations (columns), respectively; the last seven columns show the correlations between the individual subjects' relevance ratings.

Query	*Alg*	*SD*	*Mean*	*S1*	*S2*	*S3*	*S4*	*S5*	*S6*	*S7*
1	0.5678	0.1036	0.5974	0.6122	0.5585	0.4642	0.6673	0.7082	0.7029	0.4683
2	0.5831	0.0651	0.7291	0.6253	0.6631	0.7289	0.8079	0.7641	0.7825	0.7320
3	0.7034	0.1738	0.5120	0.6769	0.5908	0.5975	0.5416	0.6232	0.1852	0.3687
4	0.4262	0.0834	0.6463	0.6294	0.6220	0.6543	0.8031	0.6892	0.5475	0.5785
5	0.3858	0.1183	0.5385	0.3819	0.5617	0.6735	0.4966	0.7058	0.5134	0.4369
6	0.4374	0.1075	0.5220	0.5938	0.5319	0.5753	0.6680	0.5037	0.3403	0.4408
7	0.4097	0.1223	0.5033	0.3618	0.4281	0.6117	0.6116	0.6423	0.5103	0.3577
8	0.4595	0.1571	0.7124	0.5969	0.7907	0.8426	0.7092	0.8904	0.7269	0.4299
9	0.3269	0.1061	0.6875	0.7945	0.5692	0.6978	0.6966	0.5338	0.8226	0.6977
10	0.6277	0.1557	0.6241	0.6991	0.4461	0.7498	0.5556	0.7687	0.7508	0.3987
11	0.7422	0.0663	0.8063	0.8449	0.7307	0.8746	0.7520	0.8784	0.8322	0.7312
12	0.4983	0.0568	0.7961	0.8035	0.7724	0.8662	0.7627	0.8799	0.7550	0.7327

In order to measure the effectiveness of the matching algorithm, for each of the 12 query images we first computed the correlation coefficient between the similarity scores computed by the matching algorithm and the relevance scores obtained from the human relevance judgment study. The correlation coefficient is defined as

$$\frac{\frac{1}{n-1}\sum_{i=1}^{n}(S_i-\overline{S})(R_i-\overline{R})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(S_i-\overline{S})^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(R_i-\overline{R})^2}} \quad \text{(Eq. 1)}$$

where Si is the similarity score assigned to the ith image and Ri is its relevance score. (The Si's range from 0 to 1, and the Ri's range from 0 to 35, but they can still be correlated because the scale factor cancels in Eq. 1.) For each of the 12 query images we also computed correlation coefficients between the relevance ratings assigned by each subject and the consensus relevance scores obtained by the other six subjects. (The ratings range from 1 to 5, but here again they can still be correlated with the scores. We thus have eight correlation coefficients for each of the 12 query images; these are shown in Table 4.

We now explain how we can use these correlation coefficients to compare the performance of the matching algorithm to the performances of the individual subjects. We compute the mean and standard deviation of the seven correlation coefficients between each subject's relevance ratings and the consensus relevance score obtained from the other six subjects (the correlation coefficients in the last seven columns of Table 4). In Figure 9, for each of the 12 query images (x-axis), the vertical line indicates a range of three standard deviations from the mean, and the rectangles show the algorithm's performance compared to the human subjects' statistics. A black (white) rectangle indicates that the mean of the subjects' correlation coefficients was higher (lower) than the correlation coefficient. If the rectangle is black, its top is the mean of the subjects' correlations and its bottom is the algorithm's correlation; if it is white, the reverse is true.

For query images #9 and #12, the algorithm performed more poorly than the human subjects; its correlation was at least three standard deviations lower than the mean of the subjects' correlations. For query image #3, the algorithm's correlation was higher than the mean of the subjects' correlations. For the other nine images, the algorithm's correlation was lower than the mean of the subjects' correlations, but its performance fell within a three standard deviation range of the mean of their performances.

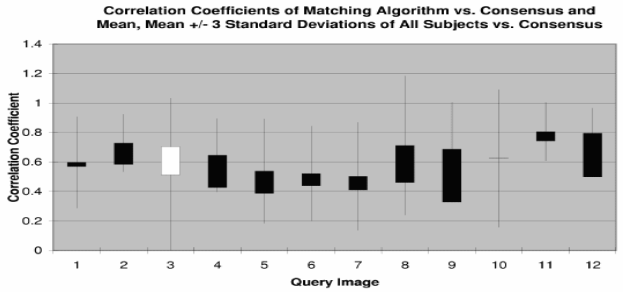

Figure 9: Comparison of the correlation coefficients in Table 4.

One of the reasons for the relatively poor performance of the algorithm is that, as shown in Figure 8, for 10 of the 12 query images less than 20% of the images received non-zero relevance ratings from at least one subject. The algorithm on the other hand, gave non-zero scores to almost all the images. This resulted in generally lower correlations between the subjects' scores and the algorithm's scores. Better results are obtained if we compute the correlations only for those images that received non-zero ratings from at least one subject. The distributions of relevance scores for these images are shown in Figure 10. The corresponding similarity scores are shown in Figure 11, and

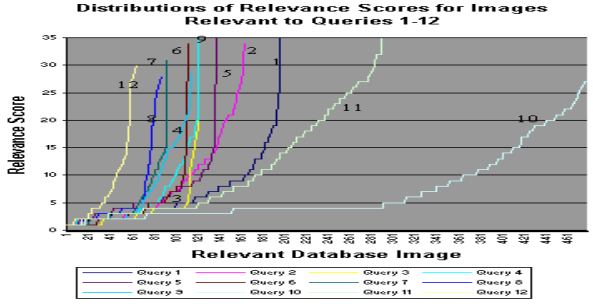

Figure 10: Distributions of relevance scores for the images that are judged relevant to query images 1-12 by at least one subject.

the correlation coefficients are shown in Table 5. Figure 12 is analogous to Figure 9, but for these correlation

coefficients only. We see that for two of the query images (#1 and #7), the algorithm's correlation was about equal to the subjects' mean correlation; for image #3, its correlation was better than the subjects' mean; and for the other nine images the correlations were all within two standard deviations of the subjects' mean (and for four of these nine images, within one standard deviation of the mean). Thus we can conclude that the algorithm's performance was comparable to the individual subjects' performance.

IV. Conclusions and Future Work

In this paper, we have described issues related to the measurement of structural similarity between document images. We defined structural similarity, and discussed the benefits of using it as a complement to content similarity for querying document image databases. We presented an approach to computing a geometrically-invariant structural similarity, and used this measure to search document image databases. Our approach supports both full image matching using query by example and sub-image matching using query by sketch. The similarity measure considers spatial and layout structure, and is computed by aggregating content area overlaps with respect to their underlying column structures. These techniques are tested within the Intelligent Document Image Retrieval (IDIR) System, and the experimental test results show that layout structure-based retrieval reflects human similarity assessments. We found that the recall performance of our proposed structural similarity retrieval method improves as the level of consensus among humans increases as well as ranking of similarity judgments increases.

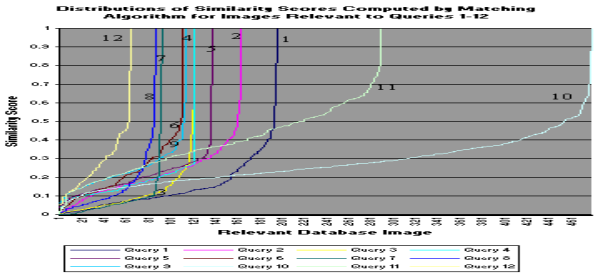

Figure 11: Distributions of similarity scores for the images of Figure 6.

In this paper we have focused on retrieval based on layout structural similarity, and we have demonstrated that a structural similarity measure based on a simple set of layout structural features behaves similarly to human similarity assessments. We intend to extend structural similarity-based document image retrieval to logical structures, and to provide learning mechanisms to find sets of weights for structural features that maximize retrieval performance. We also plan to develop a document type classification techniques based on the structural similarity, and use such techniques in document image retrieval.

References

[1] J. Blomberg, L. Suchman, and R. Trigg, Reflections on a work-oriented design project, PCD '94: Proceedings of the Participatory Design Conference, pages 99--109, 1994.

[2] D. Doermann, E. Rivlin, and A. Rosenfeld, The function of documents, IJCV, pages 799--814, 1998.

[3] Doermann, D., Shin, C., Rosenfeld, A., et al., The Development of a General Framework for Intelligent Document Image Retrieval, In J.J. Hul;, & S.L. Taylor, editors, *Document Analysis Systems II – Series In Machine Perception and Artificial Intelligence,* Volume 29, pages 433-460, World Scientific, 1998.

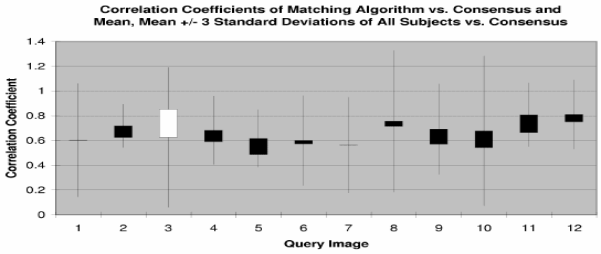

Figure 12: Comparison of the correlation coefficients in Table 5.

Table 5: Correlation coefficients. The first column ("Alg.") shows the correlations between the similarity scores and the consensus relevance scores; The second and third columns show the standard deviation and the mean of the following seven correlations (columns), respectively; the last seven columns show the correlations between the in dividual subjects' relevance ratings and the consensus relevance scores. The correlation coefficients are only for the images that are judged relevant to the query images by at least one subject.

Query	Alg.	SD	Mean	S1	S2	S3	S4	S5	S6	S7
1	0.6017	0.1529	0.6029	0.6249	0.6219	0.4044	0.7194	0.7900	0.6729	0.3870
2	0.6236	0.0583	0.7192	0.6619	0.6554	0.6860	0.7942	0.7764	0.7676	0.6927
3	0.8510	0.1885	0.6258	0.8344	0.7178	0.6326	0.6507	0.7512	0.2554	0.5383
4	0.5892	0.0924	0.6832	0.6902	0.7050	0.6643	0.8328	0.7345	0.6177	0.5381
5	0.4852	0.0778	0.6174	0.5172	0.7297	0.6460	0.6576	0.6641	0.5744	0.5324
6	0.5716	0.1210	0.6003	0.7186	0.6703	0.5345	0.7725	0.5209	0.5449	0.4405
7	0.5633	0.1286	0.5630	0.4367	0.5643	0.6189	0.6230	0.7015	0.6550	0.3418
8	0.7138	0.1909	0.7569	0.6768	0.8738	0.8891	0.7590	0.9317	0.7971	0.3709
9	0.5691	0.1222	0.6923	0.8328	0.5958	0.7159	0.6904	0.5140	0.8532	0.6441
10	0.5413	0.2020	0.6784	0.7980	0.5123	0.8476	0.6531	0.8282	0.8000	0.3092
11	0.6641	0.0858	0.8085	0.8833	0.7878	0.8839	0.8124	0.8626	0.7921	0.6375
12	0.7495	0.0930	0.8109	0.8730	0.8397	0.8754	0.8302	0.8964	0.7095	0.6524

[4] M.A. Eshera, & K.S. Fu. An image understanding system using attributed symbolic representation and inexact graph-matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:604–618, 1986.

[5] P. Herrmann, & G. Schlagetar. Retrieval of document images using layout knowledge, *Proc. International Conference on Document Analysis and Recognition*, 1993, 537–540.

[6] K. Koffka, Principles of Gestalt Psychology, Harcourt, Bryce and World, New York, 1935.

[7] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin, The qbic project: Querying images by content using color, texture and shape, SPIE 1993 Intl. Symposium on Electronic Imaging: Science and Technology, Conf. 1908, Storage and Retrieval for Image and Video Databases, February 1993.

[8] E.G.M. Petrakis, & C. Faloutsos, Similarity searching in large image databases, Technical Report CS-TR-3388, Univ. of Maryland, 1994.

[9] S. Santini and R. Jain, Similarity matching, IEEE Transactions on Pattern Analysis and Machine Intelligence, July 1995.

[10] S. Santini and R. Jain, Similarity queries in image database, In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, June 1996.

[11] Shin, C., Doermann, D., Rosenfeld, A., Classification of Document Pages Using Structure-Based Features, Special Issue on Document Analysis Systems, *International Journal on Document Analysis and Recognition*, pages 232-247, May 2001.