Content-addressable data storage by use of volume holograms

Geoffrey W. Burr, Sebastian Kobras, Holger Hanssen, and Hans Coufal

Data stored as volume holograms—optical interference patterns imprinted into a photosensitive storage material—can be accessed both by address and by content. An optical correlation-based search compares each input query against all stored records simultaneously, a massively parallel but inherently noisy analog process. With data encoding and signal postprocessing we demonstrate a holographic content-addressable data-storage system that searches digital data with high search fidelity. © 1999 Optical Society of America

OCIS codes: 200.4540, 210.2860.

1. Introduction

Volume holography has generated widespread recent interest as a possible next-generation storage technology.[1–3] Since digital data are optically input and output as two-dimensional (2-D) pages of bright and dark pixels, and are stored throughout the volume of a photosensitive storage medium, holography can offer both fast parallel access and high storage density. An added feature is content addressability,[4,5] implemented by correlation of an input data page against all of the stored pages simultaneously. If each hologram represents a database record and the input page encodes a user query, then the entire memory is searched in parallel—much faster than a conventional, software-based sequential search. Previous research has used data pages encoded to optimize the sequential, address-based holographic readout,[5] which has tended to limit the fidelity of the parallel associative searches.[6,7] Here we concentrate on maximizing the performance of the parallel search operation through data encoding and simple postprocessing. Two important steps toward an implementable content-addressable data-storage technology are made: We introduce novel encoding schemes for fuzzy database searches (to find records that are *similar* to a search argument), and we present the first, to our knowledge, demonstration of high search fidelity (using an all-holographic search-and-retrieve engine that operates on a small feature-space multimedia database).

The authors are with the IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120. G Burr's e-mail address is burr@almaden.ibm.com.

Received 17 March 1999; revised manuscript received 3 August 1999.

0003-6935/99/326779-06$15.00/0

2. Holographic Data Storage and Correlation

Figure 1 shows the three modes of operation in a holographic data-storage system: storage, address-based retrieval, and content-addressable searching. For storage, two coherent laser beams illuminate the photosensitive storage material [Fig. 1(a)]. The object beam, having passed through the pixelated spatial light modulator (SLM), contains the information to be stored. Where this beam intersects the second reference beam, a stationary interference pattern is formed, which modulates the optical properties of the storage media (such as index of refraction).

Once the hologram is recorded, either of the two beams can be used to reconstruct a copy of the other, by diffraction of a small portion of the input power off the stored interference pattern. For example, in address-based retrieval [Fig. 1(b)], the object beam can be reconstructed by illumination of the hologram with the original reference beam. Lenses image the pixelated data page onto a matched array of detector pixels, where the bright and the dark pixels can be converted back into binary data. When the hologram is stored throughout a thick storage material, then Bragg diffraction causes the strength of the reconstruction to be sensitive to changes in the angle of the reference beam. This Bragg mismatch is most sensitive to angle changes in the plane formed by the object beam and the reference beam [Fig. 1(a), the horizontal plane]. We can store and independently address multiple data pages merely by steering the

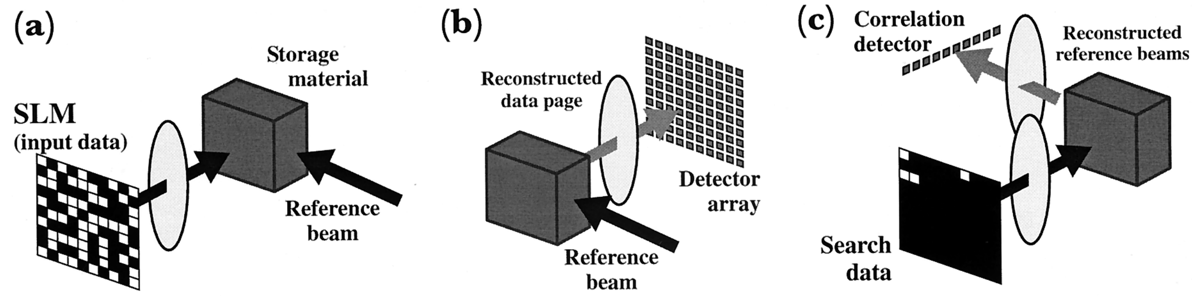

Fig. 1. Holographic data-storage system. (a) Two coherent beams, one carrying a spatial page of information, interfere within a photosensitive material to record a hologram. (b) Illuminating the hologram with the reference beam reconstructs a weak copy of the original information-bearing beam for capture with a detector array. (c) Illuminating the hologram with a new page of information reconstructs all the reference beams, computing in parallel the correlation between the search data and each of the stored pages.

reference beam (angle multiplexing). As many as 10,000 holograms have been superimposed in the same $\sim$1-cm^3 volume in this way.[8] In addition to high storage density, holographic data storage can also provide fast parallel readout. Since each data page can contain as many as 1 million pixels,[9] a readout rate of 1000 pages/s leads to an output data stream of 1 Gbit/s.

These same volume holograms, upon illumination with the object beam [Fig. 1(c)], reconstruct all the angle-multiplexed reference beams that were used to record data pages into the volume. The amount of power diffracted into each output beam is proportional to the correlation between the input data page (being displayed on the SLM), and the stored data page (recorded with that particular reference beam). Each output beam can be focused onto a detector array to form a correlation peak. Because each of the system's lenses performs a 2-D Fourier transform in spatial coordinates,[10] the optical system is essentially multiplying the Fourier transforms of the two data pages and then taking the Fourier transform of this product. The role of the hologram is to hold the complex conjugate of the Fourier transform of the stored page for multiplication by the input data and to deflect the correlation output away from the bright transmitted object beam. With a thin hologram it is possible to obtain the entire 2-D correlation function, but only for a few stored pages. This procedure has been used extensively for biometrics and target recognition.[11]

With thick holograms, the Bragg-mismatch that allows for angle multiplexing of multiple stored pages also reduces each hologram's correlation output to a narrow, nearly vertical slice of the correlation function. This slice includes the 2-D inner product (the simple overlap) between the input page being presented to the system and the associated stored page. If the patterns that compose these pages correspond to the various data fields of a database, and if each stored page represents a data record, then the optical correlation process has just compared the entire database with the search argument simultaneously. This parallelism gives content-addressable holographic data storage an inherent speed advantage over a conventional serial search, especially for large databases. For example, if an unindexed conventional retrieve-from-disk-and-compare software-based database is limited only by a sustained hard-disk readout rate (25 Mbyte/s), a search over 1 million 1-Kbyte records would take $\sim$40 s. In comparison, with off-the-shelf, video-rate SLM and CCD technology, an appropriately designed holographic system (one in which the searching object beam illuminates 1 million holograms simultaneously) could search the same records in $\sim$30 ms—a 1200$\times$ improvement. Custom components could enable 1000 or more parallel searches per second.

3. Data Encoding for Holographic Search

For this optical correlation process to represent a database search, the spatial patterns of bright (ON), pixels on the holographic data pages must somehow represent the digital data from fixed-length database fields. For example, a 2-bit data field might be encoded by four dedicated pixels within the SLM page. To find records for which this database field has a particular value, one of these four pixels is turned ON in the input page, thus reconstructing reference beams only where the stored data page had this same pixel ON. A byte (8 bits) of data might be encoded by cascading four of these 2-bit subfields. Alternatively, subfields might be organized specifically to encode letters and numbers[5] instead of bytes. When only a few matching bits are searched for, however, only a few of the SLM pixels are ON and the weak signal is often overwhelmed by the background light scatter or the detector's thermal noise. In such a case data patterns must be encoded with blocks of perhaps 10 $\times$ 10 pixels rather than with individual pixels.

The SLM is thus divided into separate regions, each dedicated to a particular fixed-length field of the database. Horizontally, these regions can be closely spaced, because Bragg mismatch will prevent input pixels from interacting with the stored pixels of different columns. However, the volume holographic correlator is still shift invariant in the vertical direction: Input pixels can and do interact with stored pixels of different rows, resulting in strong, slightly

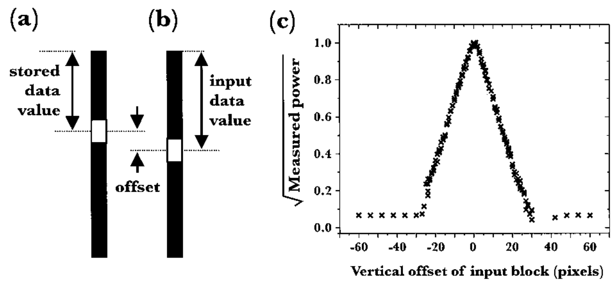

Fig. 2. Data encoding for fuzzy searching. (a) When a hologram is stored, a small block of SLM pixels are turned ON at some location within a predefined rectangular portion (slider track) of the data page. (b) For correlation readout an input query is encoded as a similar block within the same track. (c) Any offset between the two blocks causes the brightness of the correlation peak to decrease. By encoding data values with the center position of the pixel block, the holographic system can now measure the similarity between data records and the input query, thereby implementing fuzzy searching.

offset, and completely misleading correlation peaks. When several SLM pixel rows are left unused between the encoded database fields, undesired correlation peaks can be sufficiently displaced (within the output slice of the correlation function) to avoid cross talk in the measurement of the desired correlation peak (the one that carries the 2-D inner product).

The binary encoding technique described above implements an exact search through the database. By thresholding of the detected optical signal (essentially an analog quantity), the matching records can be identified. For example, during searching of a 1-byte data field encoded in four subfields, the threshold must distinguish between a full match (four of four blocks) and the closest near match (three of four). This becomes commensurately more difficult, however, when many fields are being searched simultaneously. When the threshold does not work correctly, completely unrelated records are identified as matches, because near matches between pixel block patterns do not represent near matches in encoded data value. For example, if one encodes the numerical data value 128 in the above-mentioned 1-byte data field, then near matches (three of four pixel blocks in common) will arise from records that encode many disparate values (including the data value 0). Meanwhile, the neighboring data value of 127 has no blocks in common and thus will not be considered. Gray codes can be used to guarantee that neighboring data values retain similar encodings, but they do not prevent disparate values from having equally similar encodings.

We developed a novel, to our knowledge, data-encoding method that allows for similarity or fuzzy[12] searching, by encoding only similar data values into similar pixel block patterns. As shown in Fig. 2(a), data values are encoded by the position of a block of ON pixels within a vertical track, creating a slider. For example, the data value 128 might be encoded as a pixel block of height h_s, centered within a column of 256 pixels. When data values near 128 are searched for, the partial overlap between the input slider block [Fig. 2(b)] and the stored slider block causes the resulting correlation peak to indicate the similarity between the input query and the stored data.

Using our DEMON holographic storage demonstration platform,[13] we experimentally demonstrated this fuzzy search encoding. To detect correlation peaks, a CCD camera and lens were added in the reference beam downstream from the $LiNbO_3$:Fe storage material. Figure 2(c) shows a search of a single fuzzy-encoded data field (7 pixels wide × 30 pixels tall) as the input data value approaches and then exceeds the stored value. The amplitude response (the square root of measured power as a function of the relative position of the input slider block) resembles a triangle function.[14] The correlation of identical rectangles creates the triangle; the signals add in field amplitude yet are detected in intensity, thus the square root. Since this signal is measured by the power falling on a small region of the correlation plane, only a few pixels of vertical shift invariance can contribute. In the center of the triangle, the signal added by shift invariance merely scales the slope of the triangle; at the edges of the triangle, so few pixels are involved that the correlator response is essentially unaffected by shift invariance. The choice of h_s and h_i—the vertical extent of the slider blocks on the stored and the input pages—controls the detected signal power and the search range (the height and width of the response function). Although h_s is fixed when the holograms are stored, h_i can be changed on a search-by-search basis. For further functionality, two slider blocks can be used to implement an OR (i.e., records with values <16 or >240).

With this fuzzy-encoding technique the analog nature of the optical output becomes an enabling feature instead of a drawback. However, if the cost in SLM pixels (proportional to the maximum data value) is deemed too high, data fields could be made only partly fuzzy: fuzzy representation for the low-order bits and efficient exact coding for the more significant digits. For example, the data value 128 might have its top 4 bits encoded in binary, with a much smaller slider of only 16 pixels. This trades off search flexibility for a more efficient use of the SLM area. A holographically stored database might contain a blend of fuzzy-, partly fuzzy-, and exact-coded data fields, depending on the degree of similarity matching required.

4. Search Fidelity

The capacity (both the number of stored records and the amount of data per record) that such a holographic content-addressable storage system can reach will depend on its reliability in the presence of noise. Because the optical correlation process is analog, low-overhead digital error correction is not available for the parallel search operation. One might duplicate every record in several different holograms, but this would sharply reduce capacity. A more attractive option is to use the optical parallel

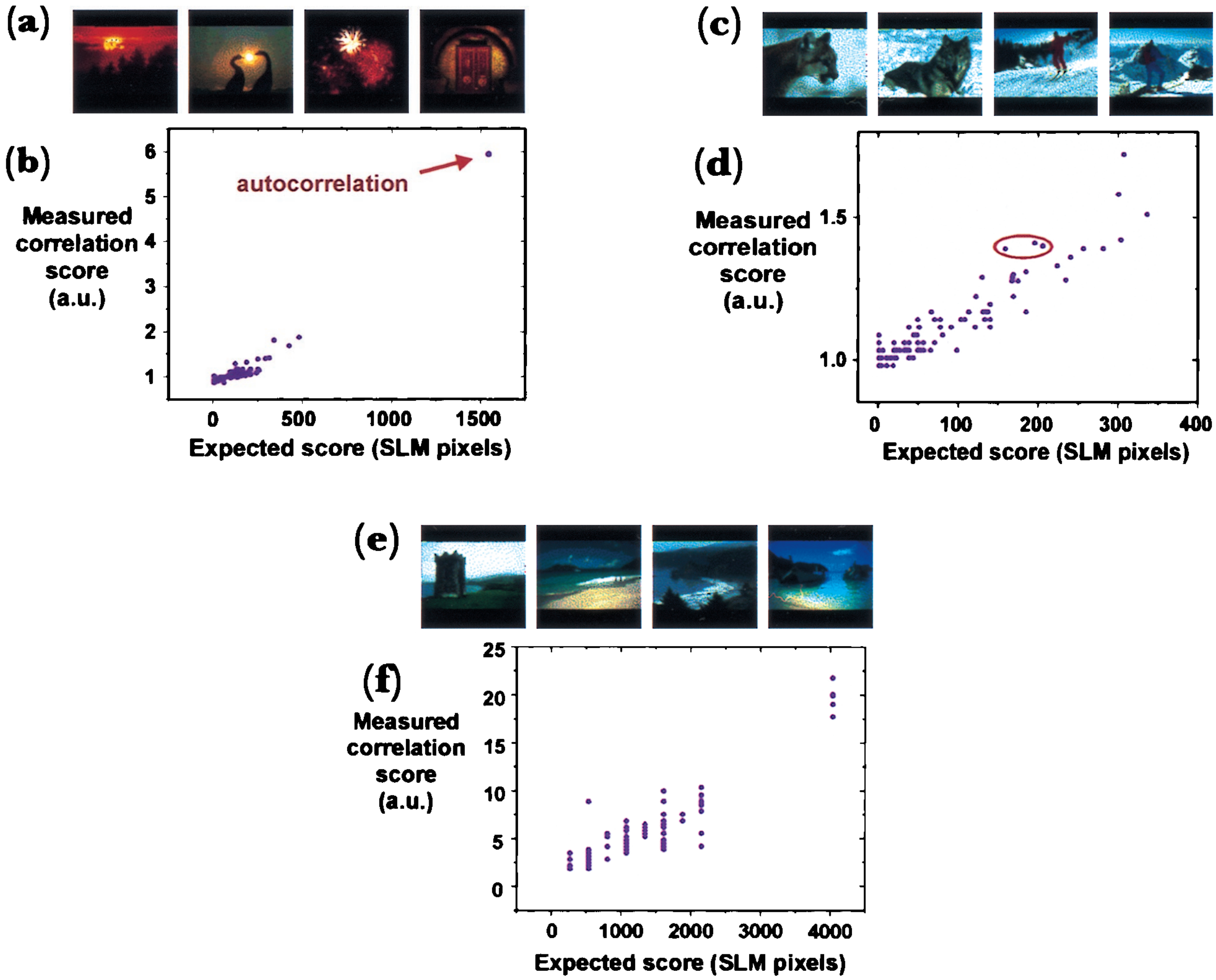

Fig. 3. Three experimental search results from an all-holographic search-and-retrieve engine, operating on a database of 100 feature vectors from the IBM QBIC image database.[15] (a) Four best images found when the search query was the color feature vector for the leftmost image. (b) Measured correlation score (ratio of the detected signal to the dark calibration value) for each of the 100 database records, as a function of the expected response (the number of SLM pixels in common between the input and each stored page). (c) Four best images found when the color sliders for 20% white and 20% light gray were input. (d) Measured versus expected correlation score. (e) Four best images found when we search for the keyword "shore," encoded into five characters with three nonbinary subfields per character. (f) Measured versus expected correlation score.

search engine as a front-end filter to a conventional sequential search engine. This retains the speed and capacity benefits of the holographic content-addressable memory while reducing the bit-error-rate burden. For example, assume the search task is to find the 10 best matches in a database of 1 million records. If the holographic system actually delivers its 100 best matches to the conventional digital search engine, then as long as the 10 best records are somewhere in this set, the optical system does not introduce any error. At the same time the number of records that the conventional search engine has to check is reduced 10,000 fold.

In addition to using the conventional search engine to relax the bit-error-rate constraints, we can make further improvements in signal-to-noise ratio. As mentioned above, the optical detection will introduce a background noise floor (from scatter or thermal noise) that sets a lower limit on acceptable signal levels. Unfortunately, the signal (detected optical energy) decreases as we increase the number of superimposed volume holograms (number of records), decrease the optical readout time (to increase search speed), or decrease the number of pixels per search field (for more data fields per record). To study these trade-offs, a diffraction model has been developed for the dependence of signal strength on the size and distribution of pixelated correlation patterns.[14,16] This model shows that defocus of the correlator (placement of the storage material away from the back focal plane of the input lens) affects the accuracy with which the detected signals measure the 2-D inner product between the input and the stored data patterns. We are currently exploring how these deterministic variations will combine with the random noise to affect the system capacity and reliability.

These deterministic signal variations can be reduced by appropriate postprocessing of the detected signals or by preprocessing of the holograms. The variations are typically introduced by hologram-to-hologram fluctuations in diffraction efficiency or by nonuniform illumination across the SLM. The former can be calibrated for each reference beam and a scaling factor applied to detected signal values[14]; the latter can be corrected during hologram recording.[17] A second calibration factor can be used to bias out the dark background signal.[14] This dark signal comes from the finite contrast of the SLM device: During a search with only a few ON pixels in the input page, the remaining pixels are not completely OFF. Although each pixel transmits only a small amount of light into the storage material, their diffracted contributions superimpose and can easily exceed the signal from a few ON pixels. Subtracting a bias factor helps but does not completely remove the problem, partly because of the deviation from the 2-D inner product described by our model[14] and partly because the bias signal often uses up most of the detector's dynamic range.

5. Experimental Demonstration

To demonstrate high-fidelity parallel searching of a holographic content-addressable memory, we stored a small multimedia database in the DEMON system.[13] Each hologram represented one record from an IBM Query-by-Image-Content (QBIC) database.[15] In the QBIC system, searches are performed across feature vectors previously extracted from the images rather than on the images themselves. Each record included several alphanumeric fields (such as image description and image number) encoded for exact searches and 64 fuzzy sliders containing the color histogram information (percentage of each given color within the associated image). To concentrate on low but nonzero color percentages, we ignored values $<1\%$, truncated values $>33\%$, and mapped the domain 1–33% to a range of 100 pixels with a square-root nonlinearity. A separate portion of the SLM page (320×240 pixels), pixel matched onto a CCD detector for conventional address-based holographic readout, was encoded with the binary GIF data for the image thumbnail (with a 6:8 modulation code[13] and 7:11 Hamming error correction code). Using an argon-ion laser at 514.5 nm, we recorded 100 holograms in an 8 mm $\times$ 15 mm $\times$ 15 mm $LiNbO_3$ crystal in the 90° geometry,[13] using a $\pm 2.5°$ span of reference angles. These reference angles were mapped onto a second CCD detector, and a 4 pixel wide $\times$ 7 pixel tall bin was assigned empirically to each focused correlation peak. After recording, the portion of the SLM not containing search data was blocked, leaving approximately 100,000 SLM pixels. A single calibration factor was then measured for each correlation bin by use of an input page with all pixels OFF.[14]

Each search, initiated by a user query, ran under computer control, including display of the appropriate patterns, detection of the correlation peaks (averaging eight successive measurements to reduce detector noise), identification of the eight highest correlation scores, mapping of correlation bins to reference beam angle, address-based recall of these eight holograms, decoding of the pixel-matched data pages, and finally display of the GIF thumbnail images on the computer monitor. Although the optical readout occupied only 0.25 s, the software-based control and decoding on the 100-MHz Pentium computer slowed the total search time, from search query to image display, to $\sim$5 s. Without the averaging, the entire optical search could be performed in a single camera frame (0.016 s). The object beam at the crystal contained approximately 20 nW for each SLM pixel turned ON (intensity contrast ratio of 167:1), whereas the readout reference beam contained $\sim$100 mW. For simplicity in this initial proof-of-principle experiment we chose to display the eight best images rather than set a threshold.

To find images on the basis of color similarity, the 64 sliders were used to input the color histogram information for the leftmost image in Fig. 3(a); the holographic search process then output the other three images as the closest-matching images. Figure 3(b) quantifies the search fidelity. As expected, the leftmost image of Fig. 3(a) correlated strongly with its own feature vector, but the system was also able to correctly identify the images with the highest cross correlation These sliders could also be used to select images by color distribution. In keeping with the QBIC algorithm, sliders for similar colors were also automatically added by the program, according to a color similarity matrix. For example, if the user asked for 30% red, the slider for 30% pink was also input. Figures 3(c) and 3(d) correspond to a search for images containing 20% white and 20% light gray. Although several images were ranked slightly higher than they deserved (red circle), the system performance was impressive, considering that the background dark signal level was twice as large as the desired signal. Note that 300 pixels corresponds to $\sim$0.1% of the 640×480 SLM—in this experiment no attempt was made to optimize the number of SLM pixels per character. In Figs. 3(e) and 3(f) the alphanumeric description field was used to search for the keyword "shore." Note that, because many characters are involved, both the expected and the measured scores are large. However, we obtained similar results for exact search arguments as small as a single character.

6. Conclusions

With the fuzzy-coding techniques and the signal-to-noise enhancements we have introduced, volume holographic content-addressable data storage is an attractive method for rapidly searching vast databases with complex queries. Areas of current investigation include implementing system architectures that support many thousands of simultaneously searched records and quantifying the tradeoffs between capacity and reliability.

We thank J. Hafner for help with the QBIC images and feature vectors; M.-P. Bernal, H. Günther, B. Lindsey, R. Macfarlane, J. Mamin, and R. Shelby for discussions; and G. Abstreiter for enabling S. Kobras's thesis study at IBM Almaden.

References and Note

1. D. Psaltis and F. Mok, "Holographic memories," Sci. Am. **273,** 70–76 (1995).
2. J. F. Heanue, M. C. Bashaw, and L. Hesselink, "Volume holographic storage and retrieval of digital data," Science **265,** 749–752 (1994).
3. J. H. Hong, I. McMichael, T. Y. Chang, W. Christian, and E. G. Paek, "Volume holographic memory systems: techniques and architectures," Opt. Eng. **34,** 2193–2203 (1995).
4. P. D. Henshaw and S. A. Lis, "Content addressable optical data-storage system," U.S. Patent 5,319,629 (7 June 1994).
5. B. J. Goertzen and P. A. Mitkas, "Volume holographic storage for large relational databases," Opt. Eng. **35,** 1847–1853 (1995).
6. B. J. Goertzen, K. G. Richling, and P. A. Mitkas, "Implementation of a volume holographic database system," Opt. Rev. **3,** 385–387 (1996).
7. P. A. Mitkas, G, A. Betzos, S. Mailis, and A. Vainos, "Characterization of associative recall in a volume holographic database system for multimedia applications," in *Advances in Optical Information Processing VIII,* D. R. Pap, ed., Proc. SPIE **3388,** 198–208 (1998).
8. G. W. Burr, F. H. Mok, and D. Psaltis, "Storage of 10,000 holograms in $LiNbO_3$:Fe," in Digest on Conference on Lasers and Electro-Optics 1994 (Optical Society of America, Washington, D.C., 1994), Paper CMB7.
9. R. M. Shelby, J. A. Hoffnagle, G. W. Burr, C. M. Jefferson, M.-P. Bernal, H. Coufal, R. K. Grygier, H. Günther, R. M. Macfarlane, and G. T. Sincerbox, "Pixel-matched holographic data storage with megabit pages," Opt. Lett. **22,** 1509–1511 (1997).
10. J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968).
11. S. H. Lee, ed., *Optical Information Processing—Fundamentals* (Springer-Verlag, Berlin, 1981).
12. L. A. Zadeh, "Fuzzy sets," Inf. Control **8,** 338–353 (1965).
13. G. W. Burr, J. Ashley, H. Coufal, R. K. Grygier, J. A. Hoffnagle, C. M. Jefferson, and B. Marcus, "Modulation coding for pixel-matched holographic data storage," Opt. Lett. **22,** 639–641 (1997).
14. S. Kobras, "Associative recall of digital data in volume holographic storage systems," Master of Science thesis (Technische Universität München, München, 1998).
15. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: the QBIC system," IEEE Comput. **28,** 23–32 (1995).
16. S. Kobras, G. W. Burr, H. Coufal, and G. Abstreiter are preparing a manuscript to be called "Optical correlation of digital data using volume holograms: diffraction analysis."
17. G. W. Burr, H. Coufal, R. K. Grygier, J. A. Hoffnagle, and C. M. Jefferson, "Noise reduction of page-oriented data storage by inverse filtering during recording, Opt. Lett. **23,** 289–291 (1998).