RELIABLE CONTENT SEARCHING WITH BEAM CONDITIONING
TECHNIQUES IN VOLUME HOLOGRAPHIC DATA STORAGE SYSTEMS

Bhargab Das*, Joby Joseph, and Kehar Singh
Photonics Group, Physics Department, Indian Institute of Technology Delhi, New Delhi-16
*Email: bhargab.das@gmail.com

Abstract: We investigate the characteristics of correlation signals in a holographic content-addressable memory in relation to the modulation codes, and the beam conditioning techniques used to obtain a homogenized object beam distribution in the recording plane. In effect, we remove the deterministic errors which result because of the presence of non-matching database records producing almost the same correlation score as the true targeted correlation score.

1. INTRODUCTION

Volume holographic data storage (VHDS) has generated widespread research interests as a possible next-generation optical storage technology. Page based data storage and the speciality of recording a large number of information carrying data pages in the same location with the help of different multiplexing techniques, results in large storage capacity, fast data transfer rates, and short random access times [1, 2]. VHDS also offers the potential for simultaneous search of an entire database by performing multiple optical correlations between the stored data pages and a search argument [3]. Such an added feature of content-addressable searching with VHDS is always beneficial considering the high storage density of 500 Gbytes in a 120 mm disk.

The concept of holographic search engine has been developed with an aim to integrating it with a digital search engine. Initially, the holographic search engine makes an approximate search of the whole database very fast, picking up a small set of nearly matching records. This small set of nearly matching records is then passed to the digital search engine to find out the best matching records with much higher accuracy. This takes full advantage of the capacity and the parallelism benefits of the holographic content-addressable memory, as also the high accuracy of the digital processor. However, care should be taken during the holographic correlation process so that non-matching records do not get included within the set of records passed to the electronic system.

2. CONTENT-ADDRESSABLE SEARCH IN A DEFOCUSED VHDS SYSTEM

In the Fourier transform (FT) geometry, when the data pages are presented in amplitude modulation mode, the recording material is illuminated very inhomogeneously because of the presence of a strong dc (zero-order) component which is not suitable for hologram recording. This dc component easily exceeds the dynamic range of the photosensitive recording material, and results in loss in terms of reconstruction fidelity of the recorded holograms [4].

This necessitates the use of object beam-conditioning techniques so that the recording material is illuminated by a homogenized or a partially-homogenized object beam. One such ingeniously simple method is to record the holograms away from the Fourier plane [5]. However, multiple optical correlations with defocused VHDS have shown system limitations. This is because of the fact that the cross-correlation peak heights are considerably stronger, and are comparable to the auto-correlation peak height. The dc-spread region gives unwanted cross-correlations, and might result into false-hits while searching a large database with partial or small search arguments.

Two different methods have been reported in the literature to remove these unwanted cross-correlation peaks. One of the methods is to perform (amplitude-recording)-(phase-searching) [6]. In this method, binary data pages are recorded in conventional amplitude mode composed of ‘ON’ and ‘OFF’ pixels. However, while searching, the search arguments are displayed in phase mode composed of ‘0’ and ‘π’ phase modulated pixels. The second method is to perform a dc-filtering operation in the FT plane during content searching operation [7]. In this technique, the data pages are in the amplitude mode both during recording and searching but we use a filter in the FT plane to block the dc-component while searching only. In both the above mentioned methods, the correlation performance for balanced binary data pages improves, with the cross-correlation peaks being significantly smaller than the auto-correlation peak. However an experimental analysis performed with data pages of various degrees of similarities has revealed some serious shortcomings of these two known methods. We discuss these issues in detail in the forthcoming sections.

3. MEASURED CORRELATION SCORE VS. DATABASE RECORD SIMILARITY FOR BALANCED BINARY DATA PAGES

Fig. 1 shows the experimental setup. Data pages
have been recorded using a laser light of wavelength 532 nm, and the recording material is a photorefractive crystal. The recording material has been positioned at a distance of 2.5 cm away from the FT plane (focal length of the FT lens is 25 cm). The input data pages are of size 512 × 512 pixels, with 2 × 2 pixels grouped together to form 1 bit of data. We have used a 90° twisted-nematic liquid-crystal (TNLC) spatial light modulator (SLM) as the input device. The polarizer at the output side of the SLM is removed and an additional quarter wave plate (QWP) and a polarizer are placed depending on the amplitude- or phase-mode of operation of the SLM. Also while performing the dc-filtering based searching, we introduce a high-pass filter that blocks only the dc peak in the FT plane. The size of the dc-filter used in the experiments is (245±10) μm.

The output of a holographic search engine is a set of r correlation scores, one for each stored hologram. This is the optical power in the correlation peak measured at the position at which each hologram’s reference beam came to focus during searching. Experiments have been performed for data pages with varying similarity from 0% to 100%. Data pages of different similarities are generated by randomly interchanging the positions of the “ON” and the “OFF” pixels. Initially we decide the size of the query data page, which is a fraction of the full data page. This is done by selecting a central region of the full data page. Then, within this fractional data page, the positions of ‘ON’ and ‘OFF’ pixels are interchanged in order to create search queries of varying similarities.

Figs. 2(a) and 2(b) show the measured correlation scores respectively for phase-based and dc-filtering based searching in relation to the similarity between the database records and the search query, considering the full data pages as the search query. High correlation scores are obtained for both the methods of searching even when the similarity between the database record and the search query tends to a very small value. It can be seen that there is an ambiguity in the correlation score when the data similarity goes below 50%. Correlation scores above and below the 50% similarity limit show almost similar characteristics. This behaviour of the correlation score can be understood with the help of the inner product arguments between the stored data page and the search query. This is due to the fact that a volume holographic searching process inherently performs a two-dimensional (2D) inner product operation.

From Figs. 2(a) and 2(b), we notice that if the holographic search engine feeds its search results to a conventional digital search engine based on a cutoff threshold applied to the measured correlation scores, there is a great probability that a non-matching record will be passed onto the digital search engine resulting into highly erroneous search outcomes. Thus appropriate measures have to be taken to remove this undesired correlation score behaviour of the phase-based and the dc-filtering based search methods in the defocussed VHDS system.

4. METHODS TO REMOVE THE ERRONEOUS CORRELATION SCORES

In the previous section, we have seen that the correlation scores behave in a confusing fashion when the similarities of the stored data pages and the query data page go below 50%. Thus if we can control the similarities among the different database records to lie above this limit, an error-free content-addressable search operation can be performed by means of the holographic search engine with defocused recording geometry. In order to do this, we need to modify the modulation codes which are used...
to create the holographic data pages from user binary data bits. A 2D constant weight block code is a modulation code which is generally used to encode user binary data bits into block of binary pixels (‘ON’ and ‘OFF’ respectively for binary data bit ‘1’ and ‘0’). In general, the code rate of a binary block code with ‘n’ block of pixels out of which ‘m’ pixels are in the ‘ON’ state, is given by

\[ r_c = \frac{1}{n} \log_2 \left( \frac{n}{m} \right) = \frac{1}{n} \log_2 \left( \frac{n!}{m!(n-m)!} \right) \]  

Balanced block codes, where the number of ‘ON’ and ‘OFF’ pixels are equal, have been used extensively for VHDS systems because such codes give the maximum code rate, and hence maximize the capacity of holographic data pages. However, balanced block codes have the special characteristics of inverse code words. This fact is depicted in Table-1 for 4-bit long code words. As can be seen, the code words in the lower half of the Table are exactly inverse of the upper half (e.g. Code word no. 6 is the inverse of Code word no. 1, Code word no. 5 is the inverse of Code word no. 2, and Code word no. 4 is the inverse of Code word no. 3). Thus while using these balanced block codes for data page coding, the similarities among the different data base records can spread out in the whole range from 0% to 100%, which is not desirable. So to create holographic data pages where the similarity between any two database records does not go below 50%, we have to discard the use of one half of the code words. Data pages constructed using this method have similarities among the different database records up to a lower limit of 50% irrespective of full or partial pages for searching.

Another possible method is the use of sparse codes for constructing the data pages where the number of ‘ON’ pixels is much smaller than ‘OFF’ pixels. A 25% sparse code (number of ‘ON’ pixels is only 25% of the total number of pixels in a block modulation code) can also perform the same desired task as mentioned earlier. Utilizing 25% sparse codes, the similarity between any two database records essentially remains above 50% (due to the nature of the code itself), irrespective of considering full or partial data pages for holographic searching. However one needs to know the maximum achievable code rate with the above mentioned two methods. Fig. 3 shows the code rates that can be achieved with the above described two methods, calculated using Eqn. (1). The code rates are plotted for different block sizes. The rejection of inverse codes in the case of balanced block codes affects severely the achievable code rate and is much lower than the corresponding code rate achieved with the 25% sparse codes. This difference is ~28% at a block size of 40. Since a higher code rate means a higher data capacity, we consider the method of 25% sparse code as the most feasible solution for removing the erroneous correlation scores.

Table 1. List of possible 4-bit balanced code words.

<table>
<thead>
<tr>
<th>No.</th>
<th>Code word</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 0 1 1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0 1 0 1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0 1 1 0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1 0 0 1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1 0 1 0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1 1 0 0</td>
<td>Inverse</td>
</tr>
</tbody>
</table>

Fig. 3 Code rate \( r \) of binary block modulation codes as a function of the block size.

Experiments on the content-addressable search have been performed for the 25% sparse data pages. The result for the dc-filtering based searching method with 25% sparse data pages is shown in Fig. 4. In this case of 25% sparse pages, the data page similarity between two randomly chosen database records does not go below 50%. This is because of the fact that, even if we interchange randomly the position of the 25% ‘ON’ pixels with ‘OFF’ pixels, there will still be 50% ‘OFF’ pixels whose positions in a data page remain unchanged. So, we can perform error-free content-searching with this method. The result shown in Fig. 4(a) confirms this point. Fig. 4(b) shows the result for the same when the size of the search query is exactly half of the stored full data pages. Here the results for the case of phase-based searching method are not shown as it shows very poor performance for sparse data pages [7] due to the reappearance of the dc peak.

In Fig. 5, we have shown results for the conventional amplitude based searching method, both for full and partial data pages. A comparison of the above experimental results suggests that the dc-filtering method together with the 25% sparse data pages can remove the shortcomings of the holographic search engine performed with the
We have shown that by constraining the database record similarities to lie above 50% using suitable block modulation code, the deterministic errors can be eliminated while content-addressable searching with a defocused VHDS system.

**5. CONCLUSION**

A volume holographic search engine can be used as a front-end filter to perform a parallel approximate search of a massive database. In order to take full advantage of the two-step searching scheme, it is necessary to remove any deterministic error arising in the holographic searching process. The experimental results shown in this paper have revealed the search limitations of the defocused VHDS system because of the appearance of erroneous correlation scores. We have suggested two different approaches to remove these erroneous correlation scores. These suggested methods create 2D holographic data pages where the database records’ similarities are restricted to be above 50% irrespective of the size of the search query. The 25% sparse modulation code shows better performance than the modified balanced modulation code. Experimental results with the dc-filtering based searching method have shown improvement in the correlation characteristics and the removal of the undesired correlation peaks thus enabling reliable content-addressable searching with the defocused VDHS.

**ACKNOWLEDGEMENT**

Bhargab Das is thankful to the University Grants Commission (UGC) India for the award of a Senior Research Fellowship.

**REFERENCES**


