Compressed Domain Texture Based Visual Information Retrieval Method for I-Frame Coded Pictures

Farzad Zargari, Member, IEEE, Mahdi Mehrabi and Mohammad Ghanbari, Fellow, IEEE

Abstract — In this paper we introduce a new compressed domain texture based visual information retrieval method. The proposed method is for the spatially predicted I-frames in the H.264 video coding standard. In I-Frame coding various prediction modes can be used for spatially predicting pixels of a block from upper or left side adjacent pixels. The selected prediction mode for a block indicates the way in which pixels of that block are related to their neighboring parts. Hence, we propose to use the histogram of the prediction modes as a texture descriptor of compressed I-frames. Since the method is based on independent I-Frame coded pictures, it can be used either for video analysis of H.264 coded videos or image retrieval of the I-frame based coded images such as advanced image coding. The simulation results indicate that this method has superior performance and lower computational load compared to an efficient realization of pixel domain texture based visual information retrieval method based on Gabor filter. Moreover, it is also robust to variations in image and coding parameters. Hence, this method is a powerful tool for visual information analysis in a wide range of applications.

Index Terms — Texture based retrieval, Compressed domain, I-frame, Prediction Mode Histogram (PMH), H.264 standard.

I. INTRODUCTION

Digital image and video libraries are expanding rapidly in recent years and effective retrieval of visual information according to the visual content is a challenging research topic. Content based retrieval (CBR) was introduced for managing and retrieval of visual information in image and video libraries. In content based image retrieval (CBIR) different features are used for image indexing and retrieval; color histogram [1], Gabor texture descriptor [2], edge histogram [3], and many other feature vectors that are listed in the MPEG-7 international standard [4] are among the features that can be used in CBIR. Video retrieval uses spatial features similar to those in still image retrieval along with temporally related features in the image sequences. In video indexing and retrieval, first the video is segmented into temporal shots, each representing an event or continuous sequence of actions. For each shot, one or more key frames are selected as the representative frame. The spatial features in the key frames along with the temporal features of the shots are used for comparing the segments in indexing and retrieval of videos.

Currently for ease of implementation of visual information on consumer devices, their processing is mainly carried out in the compressed domain. Hence, the CBR process for compressed visual information typically starts with decompression. In this way the decompression process and time will be added to the retrieval computations and time. Therefore it is advantageous to develop information retrieval algorithms to operate directly on the compressed data and avoiding full decompression of the coded image or video [5]. The compressed domain feature vectors such as those in [5-8], are introduced in the literature to achieve efficient retrieval without the need for full decompression of the visual information. Compared to the pixel domain feature vectors that extract the feature from visual content of the image or video, the compressed domain feature vectors extract the features directly from the compressed data in the coded image or video. Therefore compressed domain feature vectors depend on the visual content and also the compression techniques.

The most common video coding standards are among those in MPEG and H.26X families. These video coding standards employ hybrid coding methods including DCT block based transform and motion compensation. The MPEG and H.26X video coding standards include intra and inter frame coded pictures. The intra coded pictures (I-Frames) are coded independently and without reference to other frames. The inter frame coded pictures (P and B frames) are coded with reference to other frames using motion prediction vectors. Even though in the previous versions of these video coding standards I-pictures are coded similar to the JPEG image coding standard, in the recently introduced MPEG-4 AVC or H.264 standard they are coded using spatial prediction.

Since all these video coding standards and also the JPEG image coding standard are DCT block based, the DCT coefficients of the coded pictures are widely used to generate the compressed domain feature vectors from the compressed visual data. The compressed domain indexing and retrieval methods for DCT block based coded images and videos can be divided into two main groups. The first group includes those indexing and retrieval methods, which are specifically introduced for DCT based coded images. These methods can
be applied to the images coded by the JPEG standard and also to the I-frames in the MPEG and H.26x video coding standards, prior to MPEG-4 AVC and H.264. The works in [7-10] are among these image retrieval and indexing methods. In [7] the DCT coefficients are used to find the block or sub-block with a dominant color and the DC coefficient of it is used as the dominant color of the block or sub-block. The DC coefficients of intra coded frames in [8] are used to construct the color descriptor of compressed video in a video browsing system. The method in [9] employs the first two AC coefficients of the DCT transform to generate edge histograms similar to the edge histogram descriptors defined by MPEG-7 standard. In [10] the DCT coefficients are grouped to generate various image descriptors such as color, shape and texture descriptors.

The second group of compressed domain CBR methods for MPEG and H.26x video coding families include those indexing and retrieval methods that apply to inter-frame coded pictures. The CBR methods in [11-18] fall to this category. [11-13] extract the DCT coefficients in motion compensated pictures using the motion vector information and the DC coefficients in the I-pictures. The retrieved DC coefficients are then used for indexing various frames in the coded videos. The DC coefficients that are produced using the method in [12] are employed in [14] for scene analysis and detecting shot boundaries on compressed videos. In [15] and [16] the inter-frame prediction mode (forward, backward and skip) information besides the DCT coefficients are employed to detect cuts and abrupt scene changes in the MPEG coded videos. A frame type independent set of motion vectors (named “flow”) for each frame in the MPEG compressed videos are produced in [17]. This unified set of temporal information can be used for comparing any kind of frames in the MPEG coded videos. In [18] a hard disk drive embedded digital satellite receiver with a real time scene change detector is introduced. The detector uses the statistics of macro block types in B frames to detect scene change in compressed domain.

Since I-frames of the MPEG-4 AVC and H.264 video coding standards are coded using spatial prediction, none of the above compressed domain indexing methods for the previous video coding standards can be applied directly to the I-frames in the MPEG-4 AVC and H.264 coded videos. In this paper we propose a texture based compressed domain CBR method that can be applied directly to the compressed I-frames of MPEG-4 AVC and H.264 coded videos, and also to the single images coded by the methods such as advanced image coding (AIC) or modified advanced image coding (M-AIC) [19], which use intra frame block prediction.

The proposed texture based method for indexing and retrieval of I-frame in H.264 coded video has two important features. First, processing I-frame is independent of other frames in a group of pictures (GOP); hence the proposed descriptor can be extracted easily and rapidly from the coded video. Second, in the video coding methods, which use I-frames to code key frames [20], or compressed domain video indexing and retrieval methods, which use I-frames as the best candidate for key frames [15], the proposed method can be used for analysis of key frames in video analysis applications.

Robustness to the image and coding parameters variations is an important factor for the effectiveness of a compressed domain image retrieval method. In fact, there are proposed several methods in the literature to improve the robustness of the image retrieval methods to image parameters [21-23]. Hence, we evaluated the robustness of the proposed method to various image and coding parameters and proposed a novel method for improving the robustness of the proposed method to rotation. The performance evaluation results along with the robustness evaluations indicate that the proposed method, which is an extension of our previous works [24, 25], is a robust, low computational load, compressed domain CBIR method for I-frames, and has superior performance to other tested texture based retrieval methods. Hence, the proposed method is suitable for a wide range of compressed domain image retrieval and video analysis applications.

The rest of the paper is organized as follows. In section II the proposed texture based retrieval method is explained. Section III includes its performance evaluations and comparison with the texture based pixel domain Gabor filter method. The robustness of the proposed method to image and coding parameters are studied in sections IV and V, respectively. In section VI the performance of the proposed method for various coding parameter sets is evaluated, followed by concluding remarks in Section VII.

II. THE PROPOSED TEXTURE RETRIEVAL METHOD

In H.264 video coding standard a video can be divided into groups of pictures (GOP) with three basic types of pictures namely I, P and B [26]. I picture in a GOP is coded without reference to other pictures while P and B pictures are coded with reference to other frames. In coding each of the aforementioned picture types, first the color transform changes the color space from red, green and blue (RGB) to YCbCr. In YCbCr color space, Y is the luminance component and Cb and Cr are the chrominance components. Each color space component is divided into non-overlapping blocks and in coding I pictures, H.264 encoder uses intra prediction for each block from the outside boundary pixels of the block. Block sizes for luminance component in I-frames are either 4×4 or 16×16. H.264 encoder employs nine different prediction modes for 4×4 blocks and four prediction modes for 16×16 blocks to code I pictures. In this paper we refer to these prediction modes simply as 4×4-modes and 16×16-modes. Each 4×4-mode is a spatial prediction based on extrapolation of block boundaries in one direction, except the DC prediction mode (mode 2) that is spatial prediction based on the mean of neighboring pixels. The 4×4-modes and their corresponding directions are shown in Fig. 1.

H.264 encodes large smooth areas of picture by 16×16 blocks using four prediction modes. On the other hand, the chrominance components of I-frames are divided into non-overlapping 8×8 blocks and only four prediction modes just
Fig. 1. a) 4x4 block and the neighboring pixels used in prediction modes b) Eight spatial prediction directions in I-frame intra coding.

like the prediction modes for luminance 16×16 blocks can be used as prediction modes for chrominance blocks.

The selected prediction mode for a block in I coded picture indicates the way the pixels inside the block are related to the outside boundary pixels of the block. Fig. 2(b) shows the configuration of different prediction modes for the I-frame coded Fig. 2(a). As Fig. 2 indicates there is a close relation between the texture of original picture and the pattern of prediction modes in the I coded picture. Hence, the histogram of the prediction modes in an I-frame coded picture can be a suitable texture descriptor for the picture.

In luminance component of an I picture there are 9 prediction modes of 4×4 blocks making bins from 0 to 8 in the prediction mode histogram and 4 prediction modes for 16×16 blocks making bins 9 to 12. The 4×4-modes, which numbered as 8, 1, 6, 4, 5, 0, 7, 3 are arranged in bins 0 to 7 of the prediction mode histogram, respectively. This mapping order is based on the direction of spatial prediction mode as shown in Fig. 3. The mapping assigns the adjacent directions in the prediction modes to adjacent bins in the histogram. The remaining 4×4-mode that is DC prediction (mode 2) and all the 16×16-modes are arranged in bins 8 to 12 of the prediction mode histogram, respectively. This particular arrangement for bins of the luminance histogram is employed later in the paper to increase the robustness of the proposed method to rotation. The normalized prediction mode histogram for the luminance is defined as:

$$H_L[i] = \frac{\sum_{j=1}^{W} \text{Mode}_{L}(j)}{W}$$

(1)

There are 4 prediction modes corresponding to 8×8 blocks for chrominance components. It is worth noting that in the H.264 standard the same prediction mode is used for the corresponding blocks in Cb and Cr chrominance components. Hence, there will be only one chrominance prediction mode histogram for both of the chrominance components. The normalized prediction mode histogram for chrominance is defined as:

$$H_C[i] = \frac{\sum_{j=1}^{W_c} \text{Mode}_{C}(j)}{W_c}$$

(2)

where $H_L[i]$ is the $i^{th}$ bin; and $\text{Mode}_L(j)$ is 1, if intra coding mode of the $j^{th}$ 4×4 luminance block is equal to the prediction mode $k$ and is 0 otherwise. For 4×4-modes, Table I determines the corresponding value of $k$ for each $i$ value, and for 16×16-modes, $k$ is equal to $i$. $W$ is equal to the number of 4×4 blocks in the picture and is calculated as (picture height × picture width)/16.

<table>
<thead>
<tr>
<th>I</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>8</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>7</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

There are 4 prediction modes corresponding to 8×8 blocks for chrominance components. It is worth noting that in the H.264 standard the same prediction mode is used for the corresponding blocks in Cb and Cr chrominance components. Hence, there will be only one chrominance prediction mode histogram for both of the chrominance components. The normalized prediction mode histogram for chrominance is defined as:

$$H_C[i] = \frac{\sum_{j=1}^{W_c} \text{Mode}_{C}(j)}{W_c}$$

(2)

where $H_C[i]$ is the $i^{th}$ bin in the range of [0, ..., 3] and $\text{Mode}_{C}(j)$ is 1 if intra coding mode of the $j^{th}$ 8×8 chrominance block is equal to $i$, and is 0 otherwise. $W_c$ is equal to the number of 8×8 blocks in the picture and is calculated as (picture chrominance component height × picture chrominance component width)/64.

The similarity of luminance and chrominance prediction mode histograms, are used to measure the texture similarity between two I-frames. This similarity is calculated by using histogram intersection. We use the sum of bins in the histogram intersection of two normalized histograms as the similarity metric between luminance prediction mode histograms and chrominance prediction mode histograms of two I-frames [1]. The similarity metric between two $k$ bins
normalized histogram \( H1 \) and \( H2 \) is defined as:

\[
S = \sum_{i=0}^{r} \{ \min(H1[i], H2[i]) \} \quad (3)
\]

where \( H1 \cap H2 \) is the intersection of \( H1 \) and \( H2 \), \( \min(x,y) \) is the minimum of \( x \) and \( y \) and \( H1[i] \) and \( H2[i] \) are the \( i \)th bins in histograms \( H1 \) and \( H2 \), respectively.

The average of resulted similarity for luminance and chrominance histograms is used as the final texture similarity between two images:

\[
S_A = (S_L + S_C)/2 \quad (4)
\]

where \( S_L \) is the similarity between the luminance mode histograms and \( S_C \) is the similarity between the chrominance mode histograms. \( S_A \) is the overall similarity measure between the textures of two color images. Furthermore, since \( H_L \) and \( H_C \) are normalized, \( S_L, S_C \) and \( S_A \) are also normalized and are in the range \([0, 1]\). It is worth noting that in gray scale images only \( S_L \) represents the similarity measure between the textures of two gray scale images.

In order to increase the robustness of the retrieval method to rotation, we introduce the histogram rotation technique. The histogram rotation technique is based on the fact that picture rotation will rotate the texture of picture, and changes the direction in which pixels can be predicted from neighboring pixels. Since, the adjacent bins (0 to 7) in the defined 4×4-mode histogram are assigned to adjacent intra prediction directions (Fig. 3) texture variations by image rotation can be compensated by shifting the corresponding prediction mode histogram bins (Fig. 4).

In the rotated histogram each bin is rotated \( r \) bins to the right of its position in the original histogram. As mentioned earlier just the first seven bins of the luminance prediction mode histogram are involved in the rotation. Fig. 4 depicts the histogram rotation for a sample luminance histogram. Let \( H[i] \) is the \( i \)th bin of original histogram, then the \( r \)th bin of \( r \)-bins rotated version of \( H \) say \( H(r) \) is shown by \( H(r)[i] \) and calculated as:

\[
H(r)[i] = \begin{cases} 
H[(8 + (i + r)) \mod(8)] & \text{if } i \leq 7 \\
H[i] & \text{otherwise} 
\end{cases} \quad (5)
\]

where \( r \) is an integer number in the range of 0 to 7.

Using rotated histograms, the similarity metric (3) is calculated for the luminance prediction mode histogram of the query image and all the corresponding rotated histograms for any possible value of \( r \). The resulted maximum value of similarity metric in this stage is chosen as the similarity measure \( S_{LF}' \) between luminance prediction mode histograms:

\[
S_{LF} = \max[H1 \cap H2(r)] \quad r = 0,1, ..., 7 \quad (6)
\]

In order to increase the robustness of the proposed method to rotation, we use in (4) the calculated value for \( S_{LF} \) instead of \( S_L \). Simulation results indicate that the proposed histogram rotation technique improves the robustness of the proposed method to rotation.

III. PERFORMANCE EVALUATION

The proposed texture descriptor is used in a query by example method for retrieving images from VisTex texture image database. VisTex includes 512×512 color images of natural textures such as building, tile, cloth, etc. Each image in the database was coded as I-frame using the joint video team (JVT) H.264 encoder, and the conventional settings of the encoder such as: quantization parameter equal to 26, and the prediction mode optimization for each block was selected to minimize the residual prediction errors. The prediction mode histogram for all the coded images were derived from the compressed image files.

To evaluate the performance of the proposed texture retrieval method, a sample query image set was used. This sample query image set consists of 23 images from VisTex database with highest number of relevant images in database. For each query image the first ten retrieved images in VisTex database were recorded. These images were ranked in descending order based on the similarity value. Fig. 5 shows the first six images in the retrieved list for the given query image (the most left image in each row). We also implemented two other texture based retrieval methods. The first method was the pixel domain Gabor filter [2], which is the texture descriptor accepted in MPEG-7 standard [4]. The parameters of the filter are: mask size equal to 13×13, 24 filters with 6 orientations and 4 scales. These parameters are the optimal parameter set for the Gabor filter in image retrieval applications [27]. Using these parameters, there will be 24 output filtered images for each image and the average and standard deviation of these images make a feature vector of 48 elements. We used the similarity metric in [27] for comparing the feature vectors of Gabor filters. Since Gabor filters are only applied to the luminance component of images, in another retrieval method we used only the luminance prediction mode histogram (\( S_L \)) as the similarity measure.

For each query image we also found relevant images by inspecting the whole VisTex database. There were at most five relevant images for each image in the sample query image set. Each of the retrieved images, by the three retrieval methods marked as relevant or irrelevant using the images found in this step.

![Fig. 4. Rotation of a sample luminance histogram. The arrows show the direction of bin rotation in the rotated histogram.](image-url)
The performance measure of Average Normalized Modified Retrieval Rank (ANMRR) [28], which is defined by MPEG-7 research group, is used to evaluate the performance of the retrieval methods. ANMRR is a performance metric, which measures performance over all the points of rank retrieved list of a retrieval experiment and is expressed by a single number in the range of [0,1]. In calculating ANMRR, Each retrieval operation is assigned an NMRR, the Normalized Modified Retrieval Rate. This is averaged over all operations in the set, to produce the ANMRR:

\[
NMRR(q) = \frac{MRR(q)}{K+0.5-\frac{1}{NG(q)}}
\]  

(7)

where \( q \) is the query picture, \( K \) is relevant rank mark and \( NG(q) \) is the number of relevant pictures of query \( q \). When, in the ranked retrieved list, all relevant pictures are retrieved in top of the list, ANMRR is 0, and it will be 1 when all the relevant pictures are ranked outside the predefined threshold in the rank retrieved list that is \( K \) in (7). The rule of thumb suggested in [27] is to set \( K \) as twice the maximum number of relevant pictures for query pictures. Since query set pictures have at most 5 relevant pictures in the database, we use twice of it as the value of \( K \) in computing ANMRR in our experiments. The ANMRR values for the tested retrieval methods are shown in Table II.

The ANMRR values in Table II indicate that performance of the proposed retrieval method when uses only the luminance prediction mode histogram is comparable with the optimal Gabor filter retrieval method and the proposed retrieval method using both luminance and chrominance prediction mode histograms outperforms the optimal Gabor filter. The proposed retrieval method has also two main advantages: First, it performs in the compressed domain and does not require full decompression of the coded visual information. Second, the proposed retrieval method has lower computational load.

To provide a measure for comparing the computational load of the proposed method and Gabor filter method, we measured the average feature extraction time of the proposed method and the Gabor filter method on a 2.26 GHz PC with 1 GB of RAM. The average feature vector extraction time for 100 images in VisTex database using the proposed method and the Gabor filter method is shown in Table III. The results in Table III indicate that the proposed method has much lower computational load compared to the Gabor filter method.

<table>
<thead>
<tr>
<th>RETRIEVAL METHOD</th>
<th>ANMRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>The proposed method using Luminance &amp; Chrominance</td>
<td>0.14</td>
</tr>
<tr>
<td>Gabor filter</td>
<td>0.19</td>
</tr>
<tr>
<td>The proposed method using only Luminance</td>
<td>0.22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>METHOD</th>
<th>TIME (SECOND)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The propose method</td>
<td>0.23</td>
</tr>
<tr>
<td>The Gabor method</td>
<td>41</td>
</tr>
</tbody>
</table>
IV. ROBUSTNESS TO IMAGE PARAMETERS

In order to evaluate the robustness of the proposed image retrieval method to image parameters, we changed the selected image parameter (e.g. resolution or intensity) in the query images and made a new query image set, while leaving the image data set intact. Then, we compared the performance of the proposed method using the new query set with the performance of the method when using the original query image set. We used ANMRR for comparing image retrieval performance, for the query image sets with different image parameters. The simulation results for various image parameters are as follows.

A. Resolution

To evaluate the robustness of the proposed method to the resolution, we changed the resolution of the query set from 512×512 to 400×400, 352×352, 320×320, and 256×256. As an example Fig. 6 shows a query image in two different resolutions 512×512 and 256×256. In this way we made four new query sets, each consisting of the coded images with one of the aforementioned resolutions. The ANMRR for the retrieval experiments using query sets with different resolution are shown in Table IV. The simulation results indicate that the performance of the proposed method decreases by decreasing the resolution in the query sets but are close to the original query set except for the last two resolutions. The low performance for the last two resolutions is due to the fact that as Fig. 6 indicates, by decreasing the resolution to a high extent, some details in the image texture may be lost.

<table>
<thead>
<tr>
<th>QUERY SET RESOLUTION</th>
<th>Original</th>
<th>400×400</th>
<th>352×352</th>
<th>320×320</th>
<th>256×256</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANMRR</td>
<td>0.14</td>
<td>0.21</td>
<td>0.23</td>
<td>0.35</td>
<td>0.46</td>
</tr>
</tbody>
</table>

B. Intensity

In the evaluation of the robustness of the proposed method to image intensity, we made a new query image set by decreasing the intensity of all the images in the original query set to the same extent. To create an image with reduced intensity we subtracted a constant value from all the RGB components of each pixel at the original image as:

\[ Y = \begin{cases} 
  X - d & \text{if } (X - d) \geq 0 \\
  0 & \text{if } (X - d) < 0 
\end{cases} \]  

where \( X \) is the color component of the original image, \( d \) is the subtracted value and \( Y \) is the corresponding reduced color component.

Fig. 7 shows the resulted images for \( d \) selected as 20, 40 and 60. Table V shows the ANMRR of retrieval experiments for different intensity query sets. The results indicate that the performance of the proposed method for query sets with various reductions in intensity are very close to the original query set, as long as the reduction in the intensity does not alter the image texture, which is the case with \( d=60 \).

<table>
<thead>
<tr>
<th>INTENSITY REDUCTION IN QUERY SET</th>
<th>d=0</th>
<th>d=20</th>
<th>d=40</th>
<th>d=60</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANMRR</td>
<td>0.14</td>
<td>0.15</td>
<td>0.20</td>
<td>0.30</td>
</tr>
</tbody>
</table>

C. Rotation

Another image parameter, which tested for robustness is rotation. It is obvious that high amounts of rotation can change the texture e.g. from horizontal strips to vertical strips. Hence, a texture descriptor intrinsically cannot be robust to high amounts of rotation. We tested the robustness of the proposed method to rotation by rotating the images in the query sets and using the produced query sets to retrieve images in the original image data set. We rotated the images in the query set by 6, 12, 20, and 40 degrees to make four new query image sets. In order to compensate for the blank areas produced in the rotated images, we cropped the rotated images using a 256×256 window in the center of image. It is worth noting that we did the cropping on the images in the database using the same window, to keep the resolution consistency among query set images and database images. Fig. 8 shows a sample image and its rotated versions in various angles. We can infer from Fig. 8 that for large rotation angles such as 40 degrees, the image texture is also changed. We used the new query sets in the same way as the original query set to retrieve images from the original image data set. To evaluate the improvement of
robustness to rotation by using the histogram rotation technique, two types of retrieval experiments were conducted. First, we used the proposed method without using the histogram rotation technique and then we used the histogram rotation technique in retrieval process. Table VI shows the ANMRR of the two aforementioned retrieval experiments with rotated query sets. Table VI indicates histogram rotation technique improves the robustness of the proposed method to rotation; however, the performance falls down for large rotation angles.

**TABLE VI**
ANMRR FOR RETRIEVAL EXPERIMENTS USING ROTATED QUERY SETS

<table>
<thead>
<tr>
<th>ROTATION (DEGREE)</th>
<th>0</th>
<th>6</th>
<th>12</th>
<th>20</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANMRR (WITHOUT USING HISTOGRAM ROTATION)</td>
<td>0.21</td>
<td>0.27</td>
<td>0.39</td>
<td>0.60</td>
<td>0.85</td>
</tr>
<tr>
<td>ANMRR (USING HISTOGRAM ROTATION)</td>
<td>0.21</td>
<td>0.21</td>
<td>0.35</td>
<td>0.46</td>
<td>0.54</td>
</tr>
</tbody>
</table>

V. ROBUSTNESS TO CODING PARAMETERS

In the compressed domain image retrieval and indexing, the robustness to coding parameters is as important as robustness to image parameters, because the value of compressed domain parameters, which are used for indexing and retrieval, may change by variation in the coding parameters. In this section, we study the robustness of the proposed I-frame compressed domain texture retrieval method to two coding parameters that may affect the intra prediction mode histogram of I-frame coded images. These coding parameters are the quantization parameter and the rate-distortion optimization method for the intra prediction. The simulation results for evaluating the robustness of the proposed method to these coding parameters are as follows.

A. Quantization Parameters

Quantization parameter is used for the rate control of the compressed image or video. The higher the quantization parameter the lower is the output rate and also the lower is the quality of the decoded image. In the robustness test of the proposed method to quantization parameter, we coded images in the query image set with seven different quantization parameters: 26, 28, 30, 32, 34, 36 and 38. In this way, we made seven new query sets and we used these query sets to retrieve images from the original database, which includes images coded with quantization parameter 26. It is worth noting that quantization step doubles in size for every increment of 6 in the quantization parameter. Hence, in our experiment the quantization step incremented by a factor of 4. The ANMRRs of the corresponding experiments are shown in Table VII. These results indicate that the performance of the proposed method decreases slightly by increasing the QP from 26 to 32, which is equal to doubling the quantization step.

**TABLE VII**
ANMRR FOR RETRIEVAL EXPERIMENTS USING QUERY SETS CODED WITH VARIOUS QUANTIZATION PARAMETERS

<table>
<thead>
<tr>
<th>QP</th>
<th>26</th>
<th>28</th>
<th>30</th>
<th>32</th>
<th>34</th>
<th>36</th>
<th>38</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANMRR</td>
<td>0.14</td>
<td>0.16</td>
<td>0.27</td>
<td>0.41</td>
<td>0.55</td>
<td>0.61</td>
<td></td>
</tr>
</tbody>
</table>

B. Prediction Optimization

There are two methods for intra frame prediction mode optimization in the H.264 standard. The first method selects the intra prediction mode that minimizes the residual values after compensation, and the second method selects the intra prediction mode that optimizes the rate-distortion (R-D). The user can arbitrarily select one of these coding schemes. The first method is faster but may result in higher coding rates. Since the selected optimization mode may affect the selected mode for intra frame prediction, we studied the impact of optimization mode in the performance of the proposed CBIR method. The original data set and query set include coded images with the prediction modes that minimize the residual values and we made another query set including the same images that are coded using prediction modes that optimize the R-D. We used the new query set to retrieve images from the original data set consist of images coded with the intra prediction modes minimizing the residual values. Since the quantization parameter (QP) affects the coding rate, and hence the rate-distortion algorithm, we performed these experiments with different QPs. It is worth noting that, the higher the quantization parameter, the lower are the bitrate and the quality of the query set. Table VIII shows the ANMRRs for these query sets. Table VIII indicates that the performance decreases when using the query set coded with R-D optimization for retrieval in a data set including images coded with the intra prediction modes minimizing the residual values. However, it varies slightly by changing the QP. Furthermore, its ANMRR, which is still far from 1, indicates even in this situation the proposed method is able to retrieve relevant textures.
VI. IMPACT OF CODING PARAMETERS

Another aspect in the compressed domain visual information retrieval method, besides robustness to coding parameters, is the consistency of the performance when we use different coding parameters for coding both the query set and image data set. This is important when we use an image in a coded image data set with arbitrary coding parameters as a query image for the images in the same data set. Hence, in order to evaluate the performance of the proposed method using different coding parameters, we coded both query image set and images in the data set with similar set of coding parameters. In the first set of coding parameters the conventional settings of the encoder (as described in section III) are used including quantization parameter equal to 26 and the encoder selects intra modes to minimize the residual values. In the second set of coding parameters the rate distortion optimization mode was selected instead of minimizing the residual values. In the third experiment we only changed the quantization parameter to 38, which is the worst case in Table VII, and the encoder again selects intra modes to minimize the residual values. Table IX shows the resulted ANMRRs for the three retrieval experiments using the above coding parameter sets and indicate that when the coding parameters change simultaneously for query and data set, the retrieval performance has small variations.

VII. CONCLUSION

In this paper we introduced a novel compressed domain image retrieval method for I-frame coded images of H.264 standard. This method can be applied either to I-frames of H.264 coded videos or to the images coded by the techniques such as advanced image coding (AIC) and modified advanced image coding (M-AIC) [19], which use intra frame block prediction. The proposed method uses intra frame prediction modes histogram for visual information retrieval in texture based image retrieval and video analyses applications.

Simulation results indicate that the proposed method has lower computational load and superior performance over an efficient realization of pixel domain Gabor filter texture retrieval method. Moreover, the simulations indicate that the proposed method is robust to the variation of image parameters such as resolution and intensity. We also introduced the histogram rotation technique to improve the robustness of the proposed method to rotation. Since the proposed method is a compressed domain retrieval technique, we also tested its robustness to two important coding parameters: quantization parameter and the prediction mode optimization method. Simulation results indicate that the proposed method is robust to doubling in quantization step and is able to retrieve the relevant images when using different prediction mode optimization for coding I-frames. Moreover, the performance of the proposed method for various coding parameter sets is also tested and simulation results indicate the consistent performance of the proposed method in these experiments. Hence, the proposed texture based indexing method is effective image retrieval and indexing method. Moreover, it is robust to various image and coding parameters, and it can be used for retrieval of AIC coded images and analysis of MPEG-4 AVC and H.264 coded videos in various applications.

REFERENCES


Farzad Zargari received his B.Sc. degree in Electrical Engineering from Sharif University of Technology and his M.Sc. and Ph.D. degrees in Electrical Engineering from University of Tehran, all in Tehran, Iran. He is currently a research associate in the information technology research institute of Iran Telecom Research Center (ITRC), Ministry of Telecommunications and Information Technology of Iran. His research interests include multimedia systems, image and video signal processing algorithms, and hardware implementation of image and video coding standards.

Mahdi Mehrabi received his B.Sc. degree in electronics & telecommunication engineering from Tehran University, Tehran, Iran and his M.Sc. degree in telecommunication engineering (with honors) from Shiraz University, Shiraz, Iran. He is currently a Ph.D. candidate in the Computer Engineering Department of Islamic Azad University Science and Research Branch, Tehran, Iran. He is also a lecturer at Azad university of Shiraz, Shiraz, Iran. His research interests include computer vision and image processing, multimedia information retrieval and multimedia over IP.

Mohammed Ghanbari is the professor of video networking at the university of Essex, united Kingdom and is best known for his pioneering work on two-layer video coding for ATM networks (which earned him an IEEE Fellowship in 2001), now known as SNR scalability in the standard video codecs. He has registered for eleven international patents on various aspects of video networking and was the co-recipient of A.H. Reeves prize for the best paper published in the 1995 Proceedings of IET on the theme of digital coding. He is the author of four books and his book: An Introduction to Standard Codecs, received the year 2000 best book award by IET. Prof. Ghanbari has authored or co-authored about 450 journal and conference papers, many of which have had a fundamental influence in the field of video networking. He has served as Associate Editor to IEEE Transactions on Multimedia (IEEE-T-MM from 1998-2004).