# JPEG2000 and Motion JPEG2000 Content Analysis Using Codestream Length Information

Ali Tabesh, Ali Bilgin, Karthik Krishnan, and Michael W. Marcellin
Department of Electrical and Computer Engineering
The University of Arizona, Tucson AZ 85721-0104

## Abstract

The widespread adoption of the JPEG2000 standard calls for the development of computationally efficient algorithms to analyze the content of imagery compressed using this standard. For this purpose, we propose the use of the information content (IC) of wavelet subbands, defined as the number of bytes that JPEG2000 spends to encode the subbands. The IC of subbands can be obtained from the packet headers of the JPEG2000 codestream, thereby avoiding decompressing the arithmetically encoded bitplane data. We present experimental results for two content analysis tasks; namely, image classification and scene change detection. Our results indicate that performance comparable to that of methods operating on decompressed data can be achieved, while saving computational and bandwidth resources.

## 1. Introduction

In many applications, we need to make inferences about an image or a video sequence. Information retrieval, security applications, medical diagnosis, and remote sensing are examples of such applications wherein inferential tasks such as detection, estimation, and classification are performed. In most cases, images and video sequences are compressed prior to transmission or storage. Thus, it is highly desirable to make these inferences using the compressed domain information. In video surveillance applications, the video is usually acquired, compressed, and transmitted continuously. If the received codestream can be processed to detect suspicious activity, decompression of the unnecessary portions can be avoided. Furthermore, significant bandwidth savings can be achieved if the portions of a codestream that are of interest to a client can be identified and delivered instead of the entire codestream.

The JPEG committee has recently issued a call for contributions for standardization of technologies associated with searching image libraries [1]. This new effort is referred to as *JPSearch*. Compressed-domain analysis techniques are one of the technologies that are sought by JPSearch. These techniques offer advantages over other approaches such as wavelet-domain techniques in terms of computational and bandwidth requirements.

Myriad algorithms have been developed in the past for content analysis of images and video compressed using the JPEG and MPEG standards [2]. Although codestreams produced by those standards are structurally different than JPEG2000 codestreams, the essence of some ideas can still be used. Our work was motivated by [3] wherein the number of bits spent to encode image blocks in JPEG-coded images is used for segmentation of compressed compound documents into text, graphics, half-tones, continuous tone images, and background. However, in that work the required information is not readily available in the JPEG codestream.

Little work has been done on the analysis of JPEG2000 codestreams. In [4, 5] wavelet-domain features have been proposed for image retrieval. The proposed features can be computed at compression time. However, they do not take advantage of the particular information available in JPEG2000 codestreams.

In [6], two techniques have been proposed for indexing JPEG2000-compressed images. One technique is based on the information about the significance status of wavelet coefficients. For the lowest resolution lowpass subband, the significance map of the wavelet coefficients at all bitplanes is used as an index. The histogram $h(b,r)$ of the number of significant bits at a bitplane $b$ for subbands constituting a resolution level $r$ is used as another index. This technique requires decompressing the significance passes of the codestream, if the RESTART marker, which allows identification of individual coding passes in the compressed domain, is used at compression time. Otherwise, the entire codestream must be decompressed. The second technique proposed in [6] uses as index the means and variances of the number of nonzero bitplanes in the codeblocks of each subband. This technique takes advantage of some, but not all, of the information in the JPEG2000 codestream headers for image description.

In [7], similar ideas to those presented in this paper have been proposed for image scaling and cropping for image display applications. The scaling and cropping parameters are determined such that the most important portions of the image to be displayed are retained. The importance of each image codeblock is measured by the number of bits allocated to the codeblock, which can be determined from the packet headers of the compressed codestream. In [8], other applications of the bit allocation information are outlined. The work in [8] is different than this paper in that the main application described in that work is image segmentation and that a different statistical framework is proposed.

This paper proposes the use of the information content (IC) of subbands for content analysis of JPEG2000-compressed images and video. In the next section, we present a brief overview of JPEG2000 and MJPEG2000. We then define IC and intuitively relate it to image texture characteristics. We also describe two statistical frameworks for making inferences using the IC of subbands. The first framework is suitable for use in detection tasks, while the second framework can be used for classification. Next, we present experimental results using these frameworks for event detection in video and image classification. Finally, we present a summary and conclusions.

## 2. Overview of JPEG2000

In this section, we present an overview of JPEG2000 with emphasis on concepts related to the ideas presented in this paper. A comprehensive treatment of JPEG2000 is provided in [9]. The block diagram of a representative JPEG2000 encoder is given in Figure 1. The first stage of encoding consists of (optionally) dividing the input image into non-overlapping rectangular *tiles*. For multi-component images, e.g., color images, an optional component transform can be applied to decorrelate the components. The transformed components of each tile are referred to as *tile-components*. A wavelet transform is then applied to each tile-component and the resulting wavelet subband coefficients are partitioned into small blocks called *codeblocks*. After being quantized, the wavelet coefficients in each codeblock are entropy coded independently of other codeblocks. Entropy coding is carried out via context-dependent, binary arithmetic coding of bitplanes. The bitplane coder makes three passes over each bitplane of a codeblock. These passes are referred to as *coding passes*. Finally, the encoder forms a codestream by including coding passes selected based on a desired rate-distortion criterion.

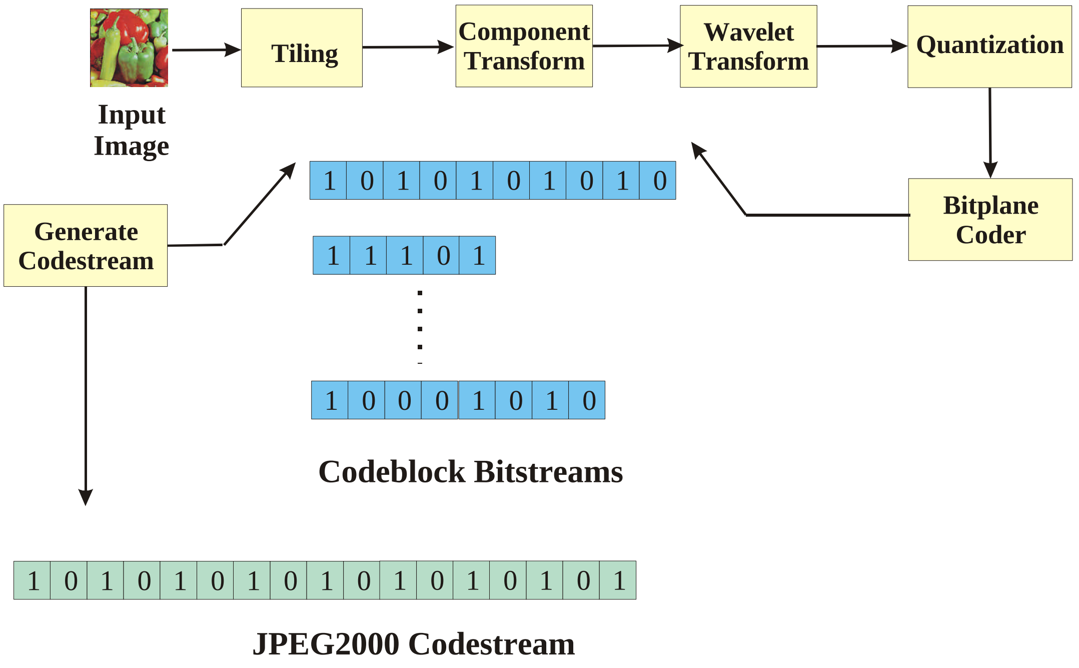

**Figure 1.** Block diagram of a representative JPEG2000 encoder.

The structure of a simple JPEG2000 codestream is given in Figure 2. This structure is explained via the notions of *precinct* and *packet*. A precinct is formed by grouping together the codeblocks that correspond to a particular spatial location at a given resolution. Compressed data from each precinct are arranged to form a packet. Each packet contains a *header* and a *body*. The packet header contains information about the contribution of each codeblock in the precinct to the packet, while the body contains compressed coding passes from the codeblocks. Packets that belong to a particular tile are grouped together to form a *tile-stream*, and tile-streams are grouped together to form the JPEG2000 codestream. Similar to packets, tile-streams are composed of a header and a body. The EOC marker indicates the end of the codestream.

MJPEG2000 is the extension of JPEG2000 for video compression [10]. Once each video frame is compressed independently using JPEG2000, the JPEG2000 codestreams may be wrapped to form a single MJPEG2000 file.

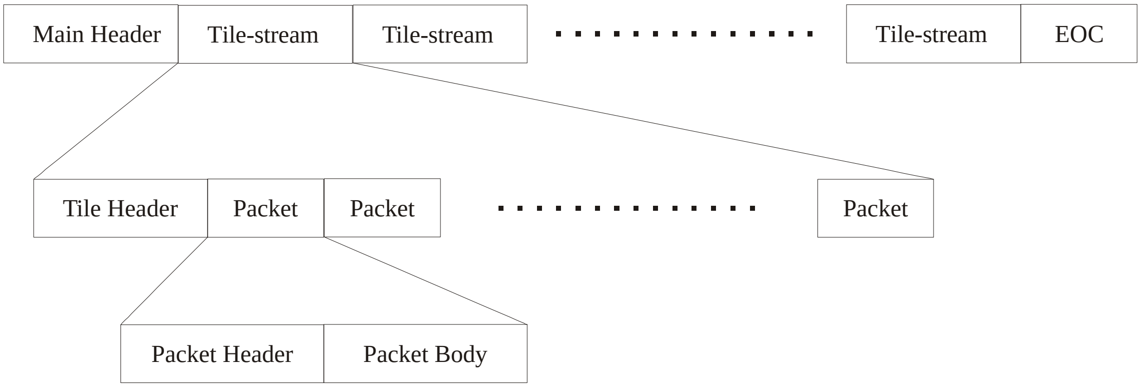

**Figure 2.** A simple JPEG2000 codestream.

## 3. Information Content of Subbands

We define the IC of a wavelet subband as the number of bytes that the JPEG2000 entropy coder spends on encoding that subband. As described in the previous section, the JPEG2000 codestream consists of a series of packets together with additional header information. Each packet header contains information about the coding passes included in the packet such as the lengths of the compressed data contributed by each codeblock within the precinct. Thus, the IC of a subband can be obtained by simply reading the headers of the packets for the corresponding resolution, and accumulating the size information for all the codeblocks within the subband. It is worth reiterating that obtaining such information from the packet headers does not require arithmetic decoding of any data. In fact, the arithmetically coded segments are contained within packet bodies and those can be completely skipped. Thus, retrieval of the IC for each subband can be performed very quickly and in a computationally efficient manner.

Intuitively, the IC of subbands conveys information about the texture characteristics of the image. Two characteristics captured by the IC are texture orientation and texture coarseness. This is demonstrated using the IC data for texture images from the Vision Texture (VisTex) database [11] with different orientation and coarseness characteristics. The IC data were obtained from the Verification Model (VM) version 9.0 [12] implementation of JPEG2000. All images were compressed at 1 bit/pixel, using $64 \times 64$ codeblocks, and three decomposition levels. Figure 3 shows the vertically oriented texture image Wood.0002 and the proportion of IC in each of its subbands relative to the total IC of the image. It can be seen that a higher proportion of image IC is in the horizontal detail subbands (0.271+0.120+0.047=0.438) than the vertical detail subbands (0.172+0.091+0.027=0.290). Figure 4 shows the coarse texture image Tile.0003 and the fine texture image Fabric.0018 and the proportion of IC for each of the four resolution levels. The finer nature of Fabric.0018 is reflected by the larger proportion of its IC in the highest resolution level, whereas the coarseness of Tile.0003 is indicated by a larger

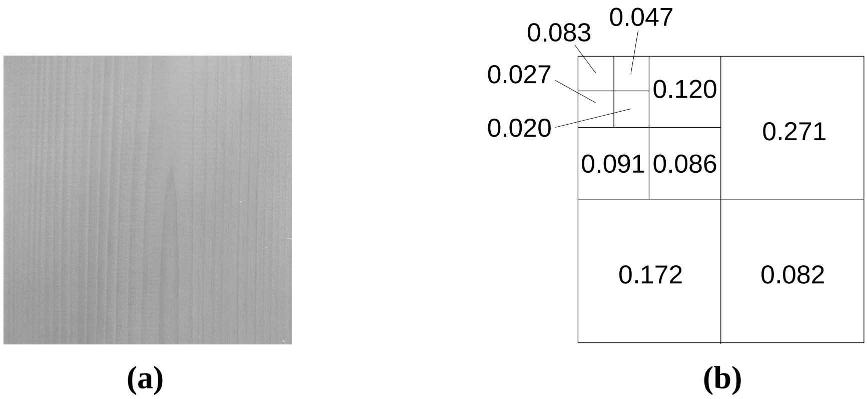

**Figure 3.** **(a)** Texture image Wood.0002; and **(b)** the proportion of IC in each of the subbands of its wavelet transform.

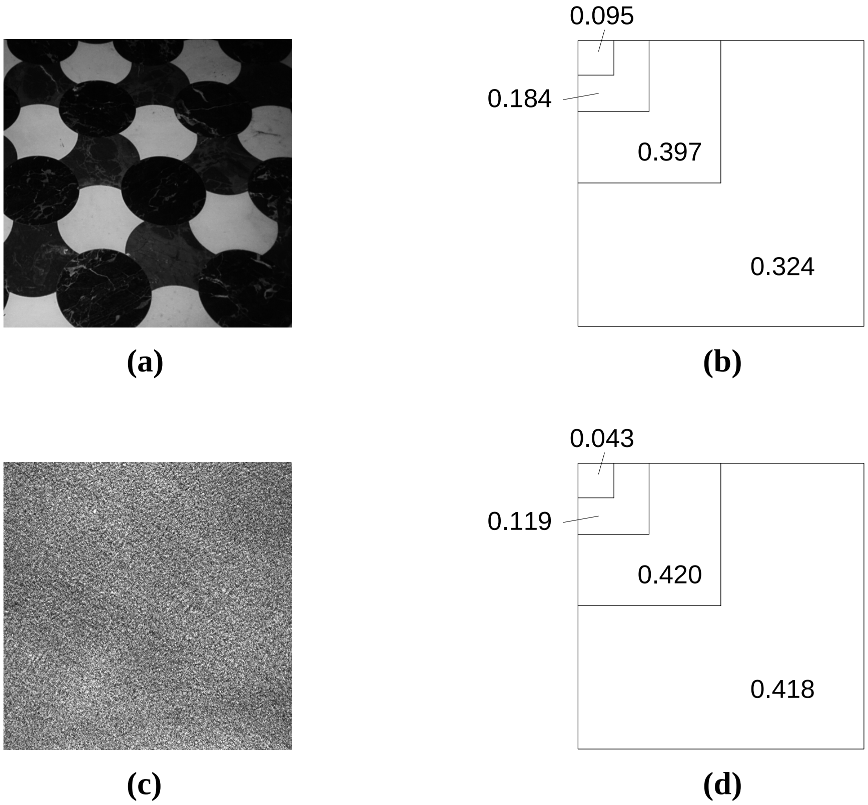

**Figure 4.** **(a)** Texture image Tile.0003; and **(b)** the proportion of IC in each of the resolution levels of its wavelet transform. **(c)** Texture image Fabric.0018; and **(d)** the proportion of IC in each of the resolution levels of its wavelet transform.

proportion of its IC in the lowest resolution level.

In the following subsections, we use the IC of video frames and images to make inferences about their texture characteristics for two applications.

### 3.1. Application to Event Detection in Video

When an abrupt change in a video sequence occurs, the texture characteristics of the frame are likely affected. This change can be captured by comparing the IC distribution of consecutive video frames. We use the $c^2$ goodness-of-fit test for this purpose. Let $n_{ij}$ denote the IC of subband $i$ of frame $j$. The two-sample $c^2$ statistic is given by [13]

$$c_j^2 = \sum_i \frac{(\sqrt{N_{j-1}/N_j}\, n_{i,j} - \sqrt{N_j/N_{j-1}}\, n_{i,j-1})^2}{n_{i,j} + n_{i,j-1}}$$

where

$$N_j = \sum\nolimits_i n_{i,j} \; .$$

If $c_j^2 > t$, where $t$ is a user-specified threshold, a scene change is declared at frame $j$. No scene change is declared otherwise. Varying $t$ controls the trade-off between false positives and false negatives.

### 3.2. Application to Texture Classification

The IC distribution of an image can be used for classification as well. This is achieved via a classification method such as the nearest neighbor (NN) algorithm [14]. Assume that every image class of interest is represented by a set of images, and the IC $n_{ij}$ for all subbands $i$ of image $j$ is known in the representative set. To classify an unlabeled image $I$, we find the "distance" between the IC $n_i^I$ of the image and that of the images in the representative set. Then, we find the representative image with the closest IC and assign its label to the unlabeled image. The distance measure between two images $I$ and $J$ used in this paper is the weighted Euclidean distance between the IC distributions given by

$$d(I,J) = \sum_{i=1}^{N} \frac{(n_i^I - n_i^J)^2}{s_i^2}$$

where $s_i^2$ denotes the variance of $n_i$ in the representative image set.

## 4. Results

We assessed the efficacy of IC distributions in two applications, namely, scene change detection and texture classification. The VM version 9.0 [12] implementation of JPEG2000 was used in the experiments.

### 4.1. Scene Change Detection in Video

We performed scene change detection on 20000 frames of a color video sequence from the movie *Batman Returns*. The frames are 640×480 pixels each. The ground truth for scene changes was established by visual examination of the frames. A total of 334 scene changes were found in the sequence. The video sequence was losslessly compressed using four wavelet transform levels, 32 × 32 codeblocks, and the reversible color transform. In evaluating all algorithms, only the Y (intensity) component of the frames was used for scene change detection. For comparison, three scene change detection methods that operate on the original frames were also implemented. The statistics used in these methods are the pixel-wise MSE between the frames, block-wise MSE between the frames using 8×8-pixel blocks, and the $c^2$ statistic between the gray-level histograms of the frames [2].

Table 1 presents the performance results for the above algorithms. Two measures have been used to characterize the detection performance: the minimum probability of error $P_e$ and the area under the receiver operating characteristic (ROC) curve (AUC). The number of false positives (NFP) and the number of false negatives (NFN) at the point on the ROC curve where $P_e$ is achieved have also been listed in the Table.

As the results indicate, the proposed algorithm achieves the second best $P_e$ and the highest AUC among the methods implemented, which is quite remarkable given the fact that it does not require the video sequence to be decompressed.

**Table 1.** Experimental results for scene change detection.

| **Method** | $P_e$ | **NFP @** $P_e$ | **NFN @** $P_e$ | **AUC** |
|---|---|---|---|---|
| **MSE** | 0.00680 | 100 | 36 | 0.9952 |
| **Block MSE** | 0.00735 | 106 | 41 | 0.9954 |
| $c^2$ | 0.01185 | 120 | 117 | 0.9935 |
| **Proposed** | 0.00730 | 83 | 63 | 0.9958 |

### 4.2. Texture Classification

We performed texture classification on a set of 480 128×128 8-bit gray-scale texture images belonging to 30 classes, with 16 images per class. The images were obtained by splitting 30 512×512 images from the VisTex database [11]. The list of texture images is given in Table 2. The images were compressed losslessly using three wavelet transform levels and 64 × 64 codeblocks. For comparison, subband energies [15] were also computed as features for classification. In our experiments, it was found that the logarithm of energy features performed better than the energy features.

The classification method for both feature sets was the NN algorithm as described in Section 3.2. The optimal feature subset for each feature set was selected by exhaustive search, i.e., by evaluating all possible feature combinations [14]. It should be noted that in practical problems with large numbers of features and training samples, a suboptimal method such as the sequential search algorithm [14] may be computationally more suitable. The quality of the features was evaluated using the probability of error $P_e$. The probability of error for both feature selection and classifier evaluation was estimated via the leave-one-out method [14]. In this method, one sample is excluded from the training set, the classifier is trained on the remaining samples, and the trained classifier is tested on the excluded sample. This operation is repeated for all samples in the training set and the number of misclassified samples is counted to estimate $P_e$. Feature selection and classification were performed using Tooldiag [16]. Table 3 summarizes the classification results using the two feature sets. As the results suggest, the IC distribution performs almost as well as the method based on the uncompressed images.

**Table 2.** Texture images used in the classification experiments.

| | | | | | |
|---|---|---|---|---|---|
| Bark.0000 | Bark.0004 | Bark.0006 | Bark.0008 | Bark.0009 | Brick.0001 |
| Brick.0004 | Brick.0005 | Fabric.0000 | Fabric.0004 | Fabric.0007 | Fabric.0009 |
| Fabric.0011 | Fabric.0013 | Fabric.0016 | Fabric.0017 | Fabric.0018 | Food.0000 |
| Food.0002 | Food.0005 | Food.0008 | Grass.0001 | Sand.0000 | Stone.0004 |
| Tile.0001 | Tile.0003 | Tile.0007 | Water.0006 | Wood.0001 | Wood.0002 |

**Table 3.** Experimental results for texture classification.

| **Feature set** | $P_e$ |
|---|---|
| **log$_2$(energy)** | 0.0354 |
| **Proposed** | 0.0375 |

## 5. Conclusions

In this paper, we have presented a new approach to texture characterization of JPEG2000-compressed images and video for content analysis. The primary advantage of the proposed approach is that it allows for rapid, compressed-domain content analysis using information obtainable from codestream headers, leading to substantial computational and bandwidth savings. Our results indicate that this approach produces accuracy comparable to that of methods operating directly on wavelet domain coefficients.

The IC of subbands can be combined with other information obtainable from packet headers, e.g., features used in [6], to further improve the analysis of JPEG2000-compressed images and video.

## References

[1] "JPSearch Scope and Requirements 1.0," ISO/IEC JTC1/SC29/WG1, Document Number 3373, July 2004.

[2] M. K. Mandal, F. Idris, and S. Panchanathan, "A critical evaluation of image and video indexing techniques in the compressed domain," *Image and Vision Computing*, vol. 17, pp. 513-529, 1999.

[3] R. L. de Queiroz and R. Eschbach, "Fast segmentation of the JPEG compressed

documents," *J. Elec. Imag.*, vol. 7, pp. 367-77, 1998.

[4] J. Bhalod, G. F. Fahmy, and S. Panchanathan, "Region based indexing in the JPEG2000 framework," in *Proc. SPIE Conf. Internet Multimedia Management Systems II*, vol. 4519, 2001, pp. 91-96.

[5] Z. Xiong and T. S. Huang, "Wavelet-based texture features can be extracted efficiently from compressed-domain for JPEG2000 coded images," in *Proc. IEEE Int. Conf. Image Proc.*, 2002, pp. I-481-I-484.

[6] M. K. Mandal and C. Liu, "Efficient image indexing techniques in the JPEG2000 domain," *J. Elec. Imag.*, vol. 13, pp. 182-187, 2004.

[7] K. Berkner, R. Neelamani, E. L. Schwartz, and M. K. Boliek, "Header-based processing of images compressed using multi-scale transforms," US Patent Application 2003/0165273, Jan 10, 2002.

[8] R. Neelamani and K. Berkner, "Adaptive representation of JPEG 2000 images using header-based processing," in *Proc. IEEE Int. Conf. Image Proc.*, 2002, pp. I-381-I-384.

[9] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Practice, and Standards*. Boston, MA: Kluwer, 2002.

[10] "Information technology -- JPEG 2000 image coding system -- Part 3: Motion JPEG 2000," ISO/IEC 15444-3, 2002.

[11] http://vismod.www.media.mit.edu/vismod/imagery/VisionTexture/.

[12] "VM 9.0 Software" ISO/IEC JTC1/SC29/WG1, Document Number 2131, April 2001.

[13] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, UK: Cambridge University Press, 1992.

[14] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.

[15] G. Van de Wouwer, P. Scheunders, and D. Van Dyck, "Statistical texture characterization from discrete wavelet representations," *IEEE Trans. Image Proc.*, vol. 8, pp. 592-598, 1999.

[16] http://www.inf.ufes.br/~thomas/home/tooldiag.html.