HVS-based medical image compression

Xie Kai a,*, Yang Jie a, Zhu Yue Min b, Li Xiao Liang a

a Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University, 200030 Shanghai, China
b CREATIS-CNRS Research Unit 5515 and INSERM Unit 630, 69621 Villeurbanne, France

Received 5 July 2004; received in revised form 8 September 2004; accepted 10 September 2004

Abstract

Introduction: With the promotion and application of digital imaging technology in the medical domain, the amount of medical images has grown rapidly. However, the commonly used compression methods cannot acquire satisfying results.

Methods: In this paper, according to the existed and stated experiments and conclusions, the lifting step approach is used for wavelet decomposition. The physical and anatomic structure of human vision is combined and the contrast sensitivity function (CSF) is introduced as the main research issue in human vision system (HVS), and then the main designing points of HVS model are presented. On the basis of multi-resolution analyses of wavelet transform, the paper applies HVS including the CSF characteristics to the inner correlation-removed transform and quantization in image and proposes a new HVS-based medical image compression model.

Results: The experiments are done on the medical images including computed tomography (CT) and magnetic resonance imaging (MRI). At the same bit rate, the performance of SPIHT, with respect to the PSNR metric, is significantly higher than that of our algorithm. But the visual quality of the SPIHT-compressed image is roughly the same as that of the image compressed with our approach. Our algorithm obtains the same visual quality at lower bit rates and the coding/decoding time is less than that of SPIHT.

Conclusions: The results show that under common objective conditions, our compression algorithm can achieve better subjective visual quality, and performs better than that of SPIHT in the aspects of compression ratios and coding/decoding time.

© 2004 Elsevier Ireland Ltd. All rights reserved.

Keywords: Medical image; Image compression; Wavelet transform; The lifting scheme; Human vision system; Contrast sensitivity function

1. Introduction

Digital technology has given a great advantage to the medical imaging area. Medical images, however, require huge amounts of memory, such as computed tomography (CT), and magnetic resonance imaging (MRI). Due to the limitations of storage and transmission bandwidth of the images, the main problem of the technology lies in how to compress a huge amount of visual data into a low-bit rate stream, because the amount of medical image data would overwhelm the storage device without an efficient compression scheme.

Recently, some researchers proposed human vision system (HVS) in the research of image compression. It can visually remove the information in most degree that human vision cannot preserve. Many solutions have been proposed for embedding a model of the human visual system in compression algorithms. While early perceptually tuned image coders were only concerned with the frequency domain behavior of the HVS [1–3], more recent and efficient coders exploit its spatial domain properties too [4–6], in order to dynamically adjust the parameters of the compression algorithm according to the local properties of the image.

Wavelet-based coders have proven ideally suited for embedding complete HVS models, due to the space–frequency localization properties of wavelet decompositions [7]. The HVS model can be embedded either in the quantization stage [4,6] or in the bit allocation stage [5]. In the former approach a perceptually tuned step is computed for quantizing each DWT coefficient; the quantized coefficient are then entropy encoded and transmitted. In the latter approach, a fixed step is used for quantizing all of the DWT coefficient, but for each
 quantized coefficient a suitable number of bits is transmitted, according to its perceptual relevance. Our algorithm takes a novel approach, which embeds an HVS model in the quantization stage of a wavelet-based coder for visually lossless compression of gray-level images. The paper is organized as follows: in Section 2 some background on the lifting scheme and HVS models is provided. Section 3 details the structure of the proposed algorithm. Finally, experimental results are presented in Section 4.

2. Background

2.1. Lifting scheme

The lifting scheme [8,9] is a new method for constructing biorthogonal wavelets. Its origins lie in a method for improving a given wavelet transform to obtain some specific properties. Subsequently, it was extended to a generic method for creating so-called second-generation wavelets. The lifting scheme is based on the interpolation to condense information. Some of its advantages as opposed to classical wavelets are the generality of the method, ease of implementation, its speed and its applicability to arbitrary length or geometries. The main difference with classical constructions is that it does not rely on the Fourier transform. Therefore, it can be used to construct wavelets in settings where translation and dilation cannot be used, such as for example, wavelets on bounded domains, on curves and surfaces or in case of irregular sampling.

It consists of three main steps: SPLIT, which subsamples the original data into odd and even sets; PREDICT, which finds the wavelet coefficients as the failure to predict the odd set based upon the even; and UPDATE, which updates the even set by using the wavelet coefficients to compute the scaling function coefficients.

2.2. Physiology of vision

Most visual properties of the HVS are not intuitive. Even when they have been characterized by psychophysical experiments, physiological evidence is the only way to understand the phenomenon completely [10,11].

The physiology of human vision includes the eyes and the retina, where vision is initiated, as well as the visual pathways and the visual cortex, where high-level perception takes place. The eyes represent the first stage of the HVS. They can be understood as a complicated camera continually in motion, allowing accommodation to different light levels and to objects at various distances. The eyes have certain optical defects such as optical blur and chromatic aberration, but normally these do not affect the rest of the processing chain.

Despite our current knowledge of the HVS, its complexity makes it impossible to construct a complete physiological model [12,13]. Some attempts have been made, but they have been restricted to models of the retina and do not account for higher-level perception. Consequently, HVS models used in image processing are usually behavioral and are based on psychophysical studies.

2.3. HVS models for imaging applications

There are also different approaches to HVS-modeling. The unifying rationale is to account for a number of psychophysical effects [14].

2.3.1. Luminance and color

The first stage in the processing chain of HVS models concerns the transformation into an adequate perceptual color space, usually based on opponent colors. After this step, the image is represented by one achromatic and two chromatic channels carrying color difference information. This stage can also take care of the so-called luminance masking or lightness non-linearity [15], the non-linear perception of luminance by the HVS. Such a non-linearity is inherent to more sophisticated color spaces like CIE L*a*b*, but needs to be added to simple linear color spaces. In compression applications, it can be considered by setting the quantization precision of the transform coefficients.

2.3.2. Multi-channel decomposition

It is widely accepted that the HVS bases its perception on multiple channels that are tuned to different ranges of spatial frequencies and orientations. Measurements of the receptive fields of simple cells in the primary visual cortex revealed that these channels exhibit approximately a dyadic structure. This behavior is well matched by a multi-resolution filter bank or a wavelet decomposition. An example for the former is the cortex transform, a flexible multi-resolution pyramid, whose filters can be adjusted within a broad range. Wavelet transforms, on the other hand, offer the advantage that they can be implemented in a computationally efficient manner by a lifting scheme.

It is believed that there are also a number of channels processing different object velocities or temporal frequencies. These include one temporal low-pass and one, possibly two, temporal band-pass mechanisms in the human visual system [16,17], which are generally referred to as sustained and transient channels, respectively.

2.3.3. Contrast and adaptation

The response of the HVS depends much less on the absolute luminance than on the relation of its local variations to the surrounding background, a property known as Weber–Fechner law [15]. Contrast is a measure of this relative variation, which is commonly used in vision models. While it is quite simple to define a contrast measure for elementary patterns, it is very difficult to model human contrast perception in complex images, because it varies with the local image content. Furthermore, the adaptation to a specific luminance level or color can influence the perceived contrast.
2.3.4. Contrast sensitivity

One of the most important issues in HVS-modeling concerns the decreasing sensitivity for higher spatial frequencies. This phenomenon is parameterized by the contrast sensitivity function (CSF). The correct modeling of the CSF is especially difficult for color images. Typically, separability between color and pattern sensitivity is assumed, so that a separate CSF for each channel of the color space needs to be determined and implemented. A chromatic CSF’s are summarized in [18], color CSF measurements are described in [19], and a detailed description for efficient CSF-modeling in combination with the wavelet decomposition can be found in [20].

The human contrast sensitivity also depends on the temporal frequency of the stimuli. Similar to the spatial CSF, the temporal CSF has a low-pass or slightly band-pass shape. The interaction between spatial and temporal frequencies can be described by spatio-temporal contrast sensitivity functions, which are commonly used in vision models for video. For easier implementation, they may be approximated by combinations of components separable in space and time.

3. The proposed algorithm

In this section the basic formulation of the algorithm is presented. Its structure resembles that of classical transform coders, and involves three sequential phases: digital transform, HVS-filters, and coding.

We adopt two schemes to improve the efficiency of the algorithm. Firstly, we use the lifting step approach for wavelet decomposition, and reduce coding/decoding time. Secondly, HVS-based image compression model is used to improve the compression ratios and the quality of recovery image.

3.1. Digital transform

The discrete wavelet transform (DWT) for two-dimensional data such as images has to be implemented indirectly. The classical structure is popularly known as Mallat’s algorithm. Although Mallat’s algorithm is widely used in image coding, it have some disadvantages. (1) In the processing of DWT, it is required to convolute with huge image data, and calculation is complex; (2) require large memory; (3) the size of image is specialized, it is not applicable to all size images.

Our problem is now to implement the above structure directly using lifting. Since a perfect reconstruction filter bank (PRFB) analysis stage can be replaced by a lifting transform block, we devised and implemented the block shown in Fig. 1.

The lifting steps scheme is particularly suitable for our purpose. First, it leads to an integer version of the DWT in a very natural way. This is of prime importance because it enables lossless coding. Then, the transform can be implemented in place, minimizing the run-time memory requirements. This can have an important impact on the computational cost when large amounts of data must be handled. Finally, it asymptotically reduces the computational complexity by a factor four. Due to the rounding operations, the integer coefficients are different from the corresponding true wavelet coefficients.

3.2. HVS-filters

The quantization noise is shaped by a filtering operation both at the encoder and the decoder side. In the following, the necessary HVS-filters for this noise shaping are introduced.

Fig. 2 shows a filterbank for an one-dimensional wavelet transformation with two levels of decomposition, where the HVS-filters are indicated by their transfer functions: $A(o)$, $B(o)$ and $C(o)$. The extension to the two-dimensional case is obvious, since all the filters are separable. For the clarity of the discussion, it is first shown what happens to the image signal at each stage (filled circles with number) of this filterbank. Then, the necessary formulae that describe the HVS-filters are given.

Stage 1 shows the input spectrum. To the right, a schematic plot of the CSF is given to indicate which part of the CSF is concerned. Stage 2 represents the output after one high-pass filtering and downsampling step. It is interesting that the corresponding subpart of the CSF needs to be applied as a high-pass filter $C(o)$. This is due to the downsampling step that shifts the portion of the initial spectrum between $\pi$ and $3/2\pi$ to the normalized frequency $\alpha$ between 0 and $\pi$. Basically, the corresponding CSF portion needs to be “mirrored” and frequency-scaled before it is applied as HVS-filter to the signal. Stage 3 shows the low-pass filtered and downsampled spectrum, which is again filtered at the second level of the filterbank. At stages 4 and 5, the concerned parts of the CSF result in two HVS-filters that have band-pass and low-pass characteristics, respectively.

If $W(f)$ denotes the measured CSF for the spatial frequency $f$, given in cycles per degree, it has to be mapped to the corresponding frequency interval in the down-sampled domain. Let $\omega_0$ be the subband specific frequency in radian relative to
the sampling rate at level $l$. Then, the transfer function of the HVS-filter, $T_{bl}^{(\omega)}$, that will be implemented in the 1D case, is given by:

$$
T_{bl}^{(\omega)} = \begin{cases} 
W \left( \frac{\pi}{2} \psi_l \right) & \text{if } b = L \\
W \left( f_A(0) - \frac{\pi}{2} \psi_l \right) & \text{if } b = H \\
W \left( \frac{\pi}{2} f_A(0) + \frac{\pi}{2} \psi_l \right) & \text{if } 1 \leq b \leq L - 1 \\
W \left( 2 - \frac{\pi}{2} \psi_l \right) & \text{if } b = 0 
\end{cases}
$$

where $b = L$ and $b = H$ correspond to the low-pass and high-pass cases. Eq. (1) represents a simple linear transformation that can be easily validated for $\omega_l = 0$ and $\omega_l = \pi$. It is in fact valid for $\omega_l \in [-\pi, \pi]$. The function $T_{bl}^{(\omega)}$ can be extended to its two-dimensional form $T_{bl}^{(\omega_l, \omega_v)}$ by introducing the horizontal and vertical frequencies $\omega_l$ and $\omega_v$.

$$
T_{bl}^{(\omega_l, \omega_v)} = \begin{cases} 
T_{vl}^{(\omega_l)} & \psi_l = 0 \\
T_{vh}^{(\omega_l)} & \psi_l = 1 \\
T_{hl}^{(\omega_l)} & \psi_l = 2 \\
T_{hh}^{(\omega_l)} & \psi_l = 3 
\end{cases}
$$

where $\psi$ is the orientation of the subband.

Now, the FIR-filter coefficients are derived by developing the Fourier series of $T_{bl}^{(\omega)}$. Since $T_{bl}^{(\omega)}$ has an even symmetry, its approximation can be expressed by:

$$
T_{bl}^{(\omega)}(n) = d_{0} + 2 \sum_{k=1}^{\infty} d_{k} \cos(\omega k) 
$$

where

$$
d_{0} = \frac{1}{\pi} \int_{0}^{\pi} T_{bl}^{(\omega)}(\omega) d\omega, d_{k} = \frac{1}{\pi} \int_{0}^{\pi} T_{bl}(\omega) \cos(k\omega) d\omega
$$

The impulse response of the HVS-filters is an FIR of length $2L + 1$, represented by the set of coefficients $\{d_2, \ldots, d_0, d_0, d_1, \ldots, d_L\}$. At the decoder side the inverse HVS-filter operation is performed. Its coefficients are computed from the inverse filter function:

$$
T^{-1}_{bl}^{(\omega)}(n) = T_{bl}^{(\omega)}(n)
$$

The choice of the filter length depends on the desired analysis/synthesis error:

$$
E_{AX}(\omega) = 1 - T_{bl}(\omega)T_{bl}^{-1}(\omega)
$$

To restrict its maximum value to a reasonable range, a filter length in the order of $L = 2$ to $L = 5$ is sufficient, depending on the steepness of the CSF subpart.

### 3.3. Coding

Due to its adaptive nature, the quantization phase zeros have high percentage of coefficients that turn out to be irrelevant to the visual quality of the compressed image. We have experimentally verified that our quantization scheme produces distribution of relevant coefficients (i.e. coefficients that are not zeroed) that is typical for DWT decompositions: many relevant coefficients reside in the coarser scale LL subband, while the remaining ones form sparse clusters concentrated in the spatial locations corresponding to edges and fine scale, high frequency details of the image. In order to efficiently encode the relevant coefficient the well-known EZW algorithm [21] is used. It must be noted that, even though an embedded coding of the image is obtained, the target bit rate should be controlled by means of the quality factor $Q$ and not by truncating the bitstream when the bit budget is exhausted, so that the visual quality of the compressed image is controlled in an HVS-driven fashion; therefore in our approach the EZW encoding proceeds until the threshold for the dominant pass reaches its final value (+1). However, this makes it impossible to set the actual bit rate with absolute precision. The embedded bitstream is finally entropy coded by means of high-performance arithmetic encoder that was presented in Ref. [22].

### 4. Experiments and results

A complete environment for testing the algorithm has been developed, based on a set of C language routines. The complete compression decompression algorithm runs in the Visual C++ environment, and calls C programs for performing EZW and arithmetic coding. From the visual environment provided by Visual C++ the user can interactively set the quality factor for each one of them and for the background.

The performance of the proposed algorithm has been evaluated by means of two sets of experiments on several medical images. The first set provides a comparison of our algorithm with well-known wavelet coders, based on rate-distortion curves; the distortion metric is the classical PSNR. However, it has been largely proved that the PSNR is an unreliable metric for measuring the visual quality of compressed images.

Therefore, a second set of experiments has been performed, in order to evaluate the actual visual quality of the compressed images.

In the first set of experiments we compare our algorithm with the well-known SPIHT coder. SPIHT was chosen due to its high performance, its widespread diffusion and the availability of an official software implementation, that allows to perform rigorous comparisons.

In Fig. 3 the rate distortion curves are reported for medical images “CT_Head” and “MRI_Head”. The performance of SPIHT, with respect to the PSNR metric, is significantly higher than that of our algorithm. The poor PSNR performance of HVS-based algorithms is due to the fact that they minimize an HVS-related metric that correlates better than the PSNR with the visual quality of the compressed image that is actually perceived by human observers.
Fig. 3. Comparison of the rate-distortion performance between the proposed algorithm and SPIHT.

Fig. 4. Despite the lower PSNR performance, the proposed method obtains, at the same compression ratios, visual qualities comparable with SPIHT.

Fig. 5. The proposed method obtains the same visual quality at higher compression ratios.
A more reliable comparison of our algorithm with SPIHT is presented in Fig. 4. It can be seen that, at the same bit rate, the visual quality of the SPIHT-compressed image is roughly the same as that of the image compressed with our approach, even if there is a big difference among the PSNR values. A visual quality comparison between our approach and SPIHT algorithm is presented in Fig. 5. Although the PSNR performance of the two algorithms is similar, our algorithm obtains the same visual quality at lower bit rates, by exploiting spatially varying characteristics of the image. The two images in Fig. 5 are almost identical, but the compression ratios are 17.12 and 15.06 for our algorithm and for SPIHT, respectively.

In Fig. 6 the compression gain provided by our algorithm over SPIHT is plotted for medical images “MRI Head” and “CT Head” for a wide range of values of the quality factor. It can be seen that our algorithm constantly outperforms SPIHT by introducing the same perceptual distortion (measured by the quality factor) at higher compression ratios.

A comparison of the coding/decoding time of the proposed algorithm and SPIHT is presented in Fig. 7. It can be seen that, at the same bit rate, our algorithm use less coding/decoding time than SPIHT.

5. Conclusion

In this paper, we use the lifting step approach for wavelet decomposition. The lifting scheme has a number of algorithmic advantages: in-place, all calculations can be performed in-place which can be an important memory savings; efficiency, in many cases the number of floating point operations needed to compute both smooth and detail parts is reduced since sub-expressions are reused; parallelism, unrolling a wavelet transform into a wiring diagram exhibits its inherent parallelism at all scales, with single write and multiple read semantics. For these reasons, we can reduce coding/decoding time.
The physical and anatomic structure of human vision is combined and the contrast sensitivity function (CSF) is introduced as the main research issue in HVS, and then the main designing points of HVS model are presented. On the basis of multi-resolution analyses of wavelet transform, the paper applies HVS including the CSF characteristics to the inner correlation-removed transform and quantization in image and proposes a new HVS-based medical image compression model. The experiments are done on the medical images including CT and MRI. The results show that the new image compression algorithm can achieve better subjective visual quality, and performs better than that of SPIHT in the aspects of compression ratios and coding/decoding time.

Acknowledgements

This work was supported by Institute of Image Processing and Pattern Recognition, Shanghai Jiaotong University. Thanks also go to Dr. Yu.

References