An Image Steganography Implementation for JPEG-Compressed Images

Kaliappan Gopalan
Department of Electrical and Computer Engineering
Purdue University Calumet
Hammond, IN 46323, U.S.A.
Tel: +1 219 989 2685, Fax: +1 219 989 2898
E-mail: *gopalan@calumet.purdue.edu*

Abstract – **A method of implementing image steganography in a color image for applications such as covert communication via the Internet and authentication of an employee carrying a picture identification card is described. By converting the color image to a one-dimensional signal in red, green, or blue, audibly masked frequencies in the 1-D signal are determined for each segment or block. Embedding of secure or confidential data is carried out by modifying the spectral power at a pair of commonly occurring masked frequencies. Compressing the data-embedded image to standard JPEG coding enables its transmission via the Internet. Experimental results show that the technique is simple to implement and causes barely noticeable distortion in the stego image. Using an oblivious technique and a key consisting of the frequencies where the spectrum is modified, successful data retrieval even at low level noise levels and at low-loss compression has been achieved. Higher payload of hidden data can be obtained at a cost of perceptibility of embedding. Lossy JPEG compression, however, leads to low payloads.**

I. INTRODUCTION

Image steganography is concerned with hiding information in a host or cover image for covert storage and/or transmission. Information is hidden using a strong key in the areas of a given cover (host) image that are imperceptible to human vision system. By concealing information in the visually masked regions of the cover image, attempts at illegal recovery and/or tampering of the hidden data are foiled. Covert communication by image steganography requires an imperceptible embedding technique that can also accurately recover the hidden information without requiring the cover image, i.e., by an oblivious method [1].

Another key application area of image embedding is in hiding vital medical or biometric information of employees in their pictures for ready access in case of an emergency, or for secure identification [2, 3]. In these applications, biometrics-based identifying information, for example, may be hidden in the picture card of a person and the claimed identity of the card carrier can be verified by retrieving the hidden data and comparing them with the biometric data collected on the spot. A small amount of distortion in image quality in such applications may be tolerable as long as data robustness is guaranteed. Watermarking of images and video for compliance of digital rights management applications, on the other hand, requires spreading of a small amount of data. While imperceptibility is critical for covert communication, data robustness and payload are vital in personnel authentication applications with, preferably, oblivious data extraction.

We present a method of embedding data in a color image that requires a key for retrieval. Image distortion causing perceptibility of embedding is minimized at a cost of lower payload, with the embedded information spread throughout the image. This method is proposed as an extension to prior work on spectral domain audio embedding by tone addition [4] and gray level and color image embedding [5, 6], both using spectral domain modification.

II. ONE-DIMENSIONAL SPECTRAL DOMAIN EMBEDDING

Based on the results of secure embedding in the spectral domain of audio signals, the proposed technique for image embedding relies indirectly on the masking property of the human visual system (HVS). In the case of audio embedding at psychoacoustically masked audio frequencies, a two step procedure has been reported [4]. In the first, a set of auditorily masked spectral points for each segment (frame) of a given cover audio signal is determined. These frequencies for each short segment of speech depend on the just noticeable difference (JND) in hearing and a global masking threshold based on a set of critical band filters. A pair of masked frequencies that occur in most segments of a given cover speech signal are obtained using a common minimum threshold of sound pressure level (SPL) relative to the global masking threshold of each segment. Since the two frequencies in each segment have SPL below the hearing (global) threshold, a slight increase in SPL up to the global threshold level (or decrease from its current level) cannot alter hearing perception. Based on this premise, SPL at the two frequencies can be set to a known ratio in accordance with data to be embedded.

Modification of the spectrum is carried out in the second step by setting the power levels at the two masked frequencies in a known ratio for bits 1 and 0. The pair of frequencies and the power ratio of the two masked spectral components (and the frame indices, if only a selected frames carry hidden data) form the key for embedding and retrieval of data. In the case of audio embedding, average power

1-4244-0977-2/07/$25.00 © 2007 IEEE

levels set to one-tenth and one-hundredth of the segment power at masked frequencies has been observed to result in inaudible and robust hiding of information. Since spectral domain modification at the two frequencies is at relatively low power levels and it is spread across all time samples in a segment, the embedded (stego) audio is rendered imperceptible from the original (cover) audio in waveform, spectrogram and audibility. Additionally, spreading of embedding makes detection by steganalysis difficult to accomplish.

Extension of spectrum modification at selected frequencies for image embedding are summarized in the next section [5, 6]. Performance of the proposed image steganography technique in terms of payload, visibility of embedding, and data retrieval accuracy are described in the following sections.

III. EXTENSION OF 1-D SPECTRUM MODIFICATION PROCEDURE TO IMAGE STEGANOGRAPHY

Exploitation of the masking effect of the HVS to imperceptibly hide information in an image generally proceeds with the determination of psychovisual contrast or pattern masking frequencies from the discrete cosine or other transform of each block of an image [7, 8]. Instead of this two-dimensional transform based method of detecting the JND in a cover image, a simpler one-dimensional approach is used in the proposed extension procedure. In this approach, the two-dimensional intensity level of a color host image in one of the three primary colors (or, the gray level of a black-and-white host image) is converted to one-dimensional signal by appending all the rows (or columns) sequentially. (It has been shown that by treating each block of 8x8 subimage as a frame (by conversion to one-dimensional data) of 'audio' and appending all the blocks together causes noticeable distortion in the image after spectrum modification [5].

Choosing a high enough 'sampling frequency,' a pair of most commonly occurring masked frequencies in all frames (of typically 64 pixel samples each) are obtained by determining global masking threshold and setting an acceptable level of spectral density below this level for each frame [9]. (Although the choice for sampling frequency is empirical, a high value – above 10,000 Hz – gives more masked frequencies, which contribute to a stronger key for embedding and retrieval.) Spectral power levels at the selected pair of frequencies in each frame (or, for added security, in selected frames) are set by a known ratio for embedding binary values. An advantage of this conversion and embedding is that it entails less computational effort and faster detection of embedding points compared to a two-dimensional transform based procedure. Since there is no relationship between the audibility of a masked tone and the visibility of a masked pixel, however, there is no guarantee that an audibly masked frequency will result in a visibly masked two-dimensional frequency pair. Still, the availability of a set of audibly masked frequencies offers a choice for spectrum modification.

A pair of most commonly occurring masked frequencies *f1* and *f2*, in part, form the key for embedding and retrieval. Complex spectrum at each of the two frequencies is modified to attain imperceptibility of embedding. Since the two audibly masked frequencies may not be present in all the segments, raising their power levels based on the global audio masking threshold for many, but not all, of the frames may result in discernibility of embedding in the overall audio and, hence, image. To prevent this, power levels at *f1* and *f2* are set to low levels in each segment. If *X(f1)* and *X(f2)* are the spectral components at frequencies *f1* and *f2* in the original segment, the spectrum-modified (i.e., data-embedded) segment is obtained as follows.

To embed a 1:

$$X'(f1) = \alpha e^{j\theta_1} \qquad X'(f2) = \beta e^{j\theta_2} \tag{1a}$$

To embed a 0:

$$X'(f1) = \beta e^{j\theta_1} \qquad X'(f2) = \alpha e^{j\theta_2} \tag{1b}$$

where $X'(f1)$ and $X'(f2)$ are the modified spectral components at frequencies *f1* and *f2*, and θ_1 and θ_2 are the phase angles of the original spectrum at *f1* and *f2*. The constants α and β are adapted based on the average power of each segment. Typically, α is larger than β so that the spectral magnitude at one of the two frequencies is higher than that at the other frequency. Both values, however, are small enough so that they are not visible in the spectrogram (histogram) of the audio (image) signal and large enough to be not lost in quantization after embedding. The values for α and β are set empirically for a given cover image.

Since the two frequencies *f1* and *f2* for an embeddable frame are in its masked region, adding or subtracting spectra at these frequencies ensures that the modification results in minimal change in 'audibility.' Also, by retaining the same phase as that in the original spectrum at *f1* and *f2*, no phase-related distortion is introduced after modification. Modified segment (frame) spectrum is transformed to time domain, quantized to the same number of levels as the cover image, and converted back to two-dimensional image.

For a color image, the choice of masked frequencies for each color can be different for added security. This difference in frequency pairs may also be warranted due to differences in luminosity levels. For efficient covert transmission or storage, the embedded stego image may be compressed using a low-loss compression scheme such as JPEG; lossy compression, however, may cause bit errors in data retrieval unless payload of covert information is reduced and replicated for error detection and correction.

Embedded information in each segment is recovered by the spectral ratio at the two (key) frequencies, *f1* and *f2* in each color. That is, the recovered bit *rb* in a segment in a selected color is given by

$$rb = \begin{cases} 1, \text{ if } \left|\dfrac{X_r(f1)}{X_r(f2)}\right| > b1 \\ 0, \text{ if } \left|\dfrac{X_r(f2)}{X_r(f1)}\right| > b2 \\ -1, \text{ else} \end{cases} \qquad (2)$$

where $X_r(f)$ is the spectral component of the embedded and quantized frame at cyclic frequency *f*, and *b1* and *b2* are set empirically. If a segment is left unembedded for added security or when the data size is smaller than the available capacity, spectral levels at the two frequencies are set equal; this corresponds to a retrieved 'bit' of −1 under niose-free conditions; with noise in transmission, the ratios can be set with a small threshold based on estimated noise level.

The key for embedding and retrieval consists of the indices of the embedded frames (rows or columns of 64 pixels of image), if only a selected frames are used to carry hidden data, and the corresponding frequency pair used to modify spectrum in each color. This key, clearly, depends on the cover (host) image. Additionally, a given cover image may use more than one pair of frequencies for added security or strong key. Both the variability and the presence of many masked points, in addition to the sampling frequency, make it harder for illegal retrieval and/or tampering of data by an exhaustive search of possible embedded frequencies.

IV. IMPLEMENTATION RESULTS AND DISCUSSION

Results of the two-step image embedding algorithm using the gray scale cameraman image (cameraman.tif) available in MATLAB showed that embedding at a pair of high frequencies, even though they were not the most commonly occurring masked frequencies, caused little noticeable distortion and zero bit error in data recovery [5]. Based on these results, the technique was applied to embedding data in one or more of the primary colors in a JPEG-encoded color image. This image (kid.jpg, shown in Fig. 1a) of 289x200x3 pixels was converted to one-dimensional signal in each of the three colors by appending all the rows of pixel values together. Using an arbitrary sampling frequency of 16,000 Hz, masked frequencies for each color were obtained. The pair of frequencies, 4875 Hz and 6250 Hz, which were in the masking set of fewer than 30 percent of the segments for all three colors, resulted in minimal distortion of embedded image with each of the three colors. For the size of 289x200 = 57800 values (pixels) in the one-dimensional signal, a maximum of 57800/64 = 903 bits can be embedded when all the frames of 64 pixels each are used.

To test the image quality with this full capacity, (a) all bits of 1, (b) all bits of 0, and (c) a random set of 903 bits, were used each for the data with the constants α and β set at a ratio of 1.5E−5 from the average power of each segment.

Use of the average segment power in setting the spectral ratio for bits 1 and 0 adaptively modifies the spectrum to low levels so that the perceptual difference in the image becomes negligible. The resulting image for (a) with the spectrum of the blue array modified is shown in Fig. 1 along with the original cover image. Using a spectral ratio of *b1 = b2 = 1* in Eq. (2), all the embedded bits were retrieved correctly from the modified and quantized image. Perceptibility of embedding, as can be seen from the figure, appears to be minimal.

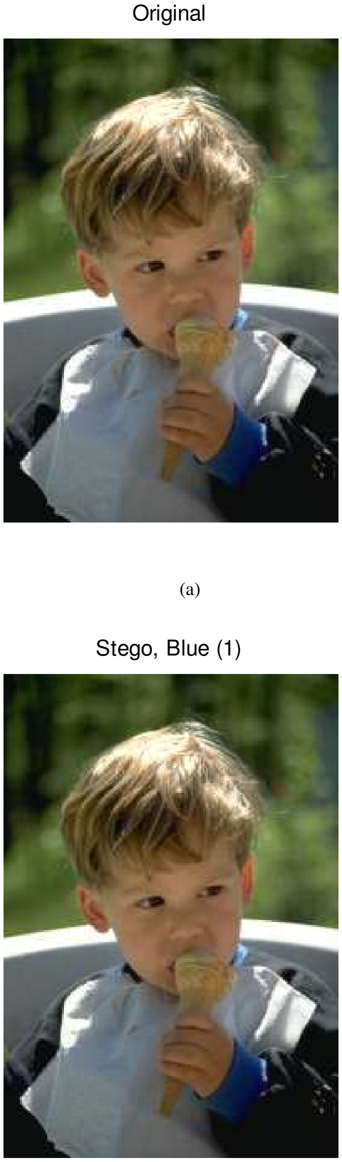

Fig. 1 (a) Original host image, and (b) Image with blue color carrying 903 bits of 1's

A pixel-by-pixel difference in the blue color array that had been changed, showed a range of [-20 22] with the blue array of the original host image, as shown in Fig. 2. Histograms of the original and modified blue part of the image, shown in Fig. 3, depict very little noticeable difference due to embedded data of all 1's. We note that in practical applications, data are rarely all 1's or all 0's. Hence, the histogram of the stego image cannot be any more different from that of the stego carrying all 1's (or all 0's). Fig. 4. shows the histogram for the case of embedding 903 bits of random data. Same results in data recovery and histograms were observed using each of red and green colors for embedding.

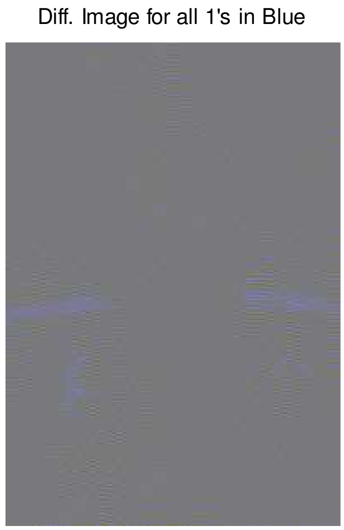

Fig. 2. Scaled difference between stego (Fig. 1b) and original (Fig. 1a)

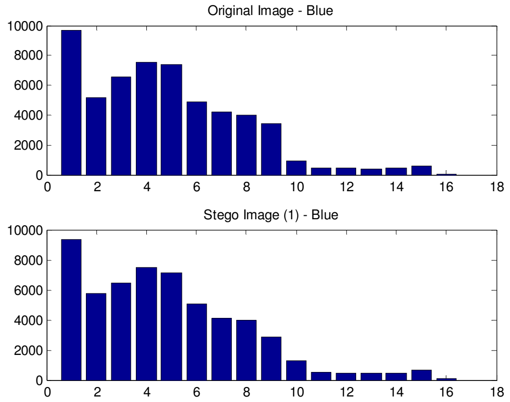

Fig. 3. Histograms of the original (top) and modified (bottom) images with blue color carrying 903 bits of all 1's

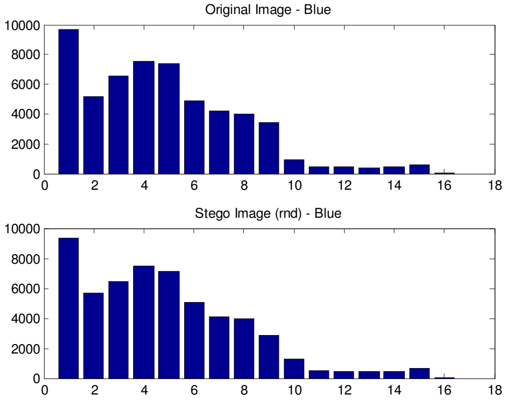

Fig. 4. Histograms of the original (top) and stego (bottom) images with blue color carrying 903 bits of random data

Extending the embedding procedure to more than one color using a different pair of frequencies for each color increases the payload threefold while making the extraction key stronger. Fig. 5 shows the cover image of Fig. 1 carrying a total of 3x903 = 2709 bits of random data. At this high payload, embedding for the choice of frequencies of 4840 Hz and 5000 Hz for red, 4880 Hz and 5040 Hz for green, and 4880 Hz and 5080 Hz for blue (corresponding to DFT indices of 122, 126, 123, 127, 123 and 128, respectively, at 256-point DFT in MATLAB at the sampling rate of 10240 Hz) is slightly noticeable. This visibility, also seen in the difference image of Fig. 6, is due to the fact that the frequencies used were not in the masked regions of *all* the frames. (The noticeable boundary between two adjacent segments of 64 pixels in the one-dimensional array is due to abrupt transition, similar to speech frames using rectangular windows. Use of a variable number of pixels in each frame to minimize this effect is under investigation.)

Difference in the histograms, however, is not as noticeable as image difference, as depicted in Figs. 7-9, unless the histograms of the original image are available for comparison. Considering that any cover image, in general, cannot have the same pair of frequencies in the masked regions of all the frames, we can minimize visibility at the expense of lower payload: if a frame does not have the selected embedding (key) frequencies in its masked regions, this frame may be excluded from spectrum modification or embedding. This exclusion strengthens the key while reducing data payload in covert communication applications employing commonly available images as hosts. If a person's identification or medial information is to be hidden in his/her picture card, however, data robustness and key-based access are more crucial to protect the information than changes in image perceptibility. With a significantly high volume of identification or other private data, therefore, it is imperative to retain the maximum payload possible. As a

Fig. 5 Stego image carrying 903 bits (random) in each color

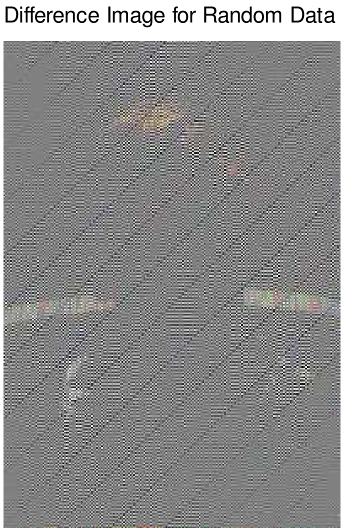

Fig. 6. Bitwise difference image between cover (Fig. 1) and stego (Fig. 5)

matter of practical importance, we note that a picture in an identification document is usually in a compressed form for storage efficiency and generally no other exactly similar image may be available for comparison. Hence, storing (or transmitting) a data-embedded image is not likely to cause noticeability of embedding.

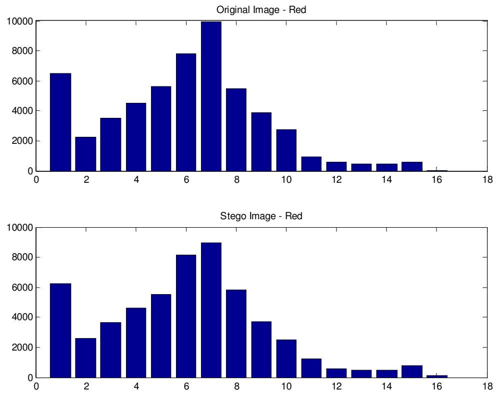

Fig. 7 Histograms of original (Fig. 1) and stego (Fig. 5) images for red

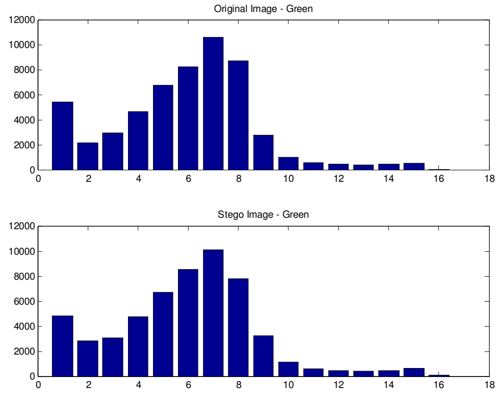

Fig. 8 Histograms of original (Fig. 1) and stego (Fig. 5) images for green

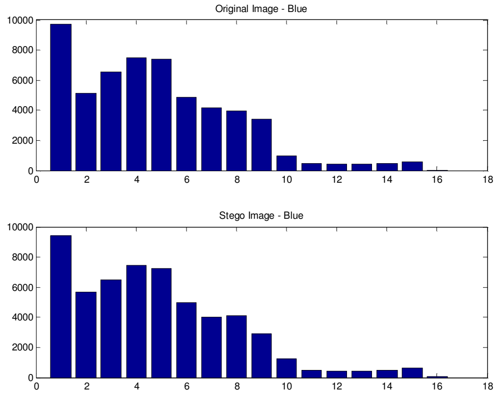

Fig. 9 Histograms of original (Fig. 1) and stego (Fig. 5) images for blue

Data recovery caused no bit errors when all three colors were employed for embedding. Successful data extraction also resulted using Eq. (2) when the embedded image was JPEG-encoded using lossless compression. Lossy compression, however, caused bit errors ranging from a few bits at 70 percent to over 30 in each color at lower quality with 903 bits of random data. Lower payload with possibly error correction and detection schemes can, therefore, be used for lower image quality. The trade off between JPEG compression, payload and bit error must be considered for covert communication and secure identification card applications. Preliminary results on the robustness of data with additive Gaussian and salt-and-pepper noise showed low bit-errors at low noise power levels [6].

V. CONCLUSION

A method of embedding data on a color image for steganography with applications in covert communication and secure identification and authentication of personnel has been proposed. By altering the one-dimensional spectrum of each segment of a cover image at two 'audibly' masked frequencies, embedding becomes barely noticeable. Availability of a choice of frequencies (for the key at which the spectrum is modified) renders a strong key and makes the hidden data impervious to unauthorized access. Another advantage of the technique is that the hidden information is extracted by an oblivious method; hence, the proposed method is suitable for transmitting embedded information using any cover image regardless of its availability at the receiver.

A key question that arises from the proposed method is the lack of correlation between audibly masked frequencies and the JND in each image frame. Another is the choice of an appropriate sampling frequency in the conversion so that an embedded image is indistinguishable from its original cover image. Since there is no relationship between the audibility of a masked tone frequency and the visibility of a masked pixel, the implicit assumption in going from one-dimensional (audible) to two-dimensional (visible) domain may not always result in imperceptible embedding. The simplicity of the proposed technique, therefore, must be weighed against these questions.

V. REFERENCES

[1] F.A. P. Petitcolas, R.J. Anderson and M.G. Kuhn, Information hiding – a survey," *Proc. IEEE*, Vol. 87, No. 7, 1999, pp. 1062-1078.

[2] M. Wu and B. Liu, "Data hiding in image and video .I. Fundamental issues and solutions," *IEEE Transactions on Image Processing*, Vol. 12, June 2003, pp. 685 – 695.

[3] M. Wu, H. Yu and B. Liu, "Data hiding in image and video .II. Fundamental issues and solutions," *IEEE Transactions on Image Processing*, Vol. 12, June 2003, pp. 696 – 705.

[4] K. Gopalan, S. Wenndt, S. Adams and D. Haddad, "Covert Speech Communication Via Cover Speech By Tone Insertion," *Proc. of the IEEE Aerospace Conference*, Big Sky, MT, Mar. 2003.

[5] K. Gopalan, "A Spectrum Modification Technique for Embedding Data in Images," *Proc. of the IEEE 38th Southeastern Symposium on System Theory*, Cookeville, TN, Mar. 2006, pp. 231-234.

[6] K. Gopalan, "A One-Dimensional Technique for Embedding Data in A JPEG Color Image," *Proc. of the IEEE Asia Pacific Conference on Circuits and Systems (APCCAS 2006)*, Singapore, Dec. 2006,

[7] B. Wolfgang, C.I. Podilchuk, and E.J. Delp, "Perceptual watermarks for digital images and video," *Proc. IEEE*, Volume: 87, Issue: 7, July 1999, pp.1108 – 1126.

[8] W. Bender, D. Gruhl, N. Morimoto and A.Lu, "Techniques for data hiding," *IBM Systems Journal*, Vol. 35, Nos. 3 & 4, 1996, pp. 313-336.

[9] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE*, vol. 88, No. 4, Apr. 2000, pp. 451-513.