Journal of Computer Science 9 (7): 928-934, 2013
ISSN: 1549-3636

doi:10.3844/jcssp.2013.928.934 Published Online 9 (7) 2013 (http://www.thescipub.com/jcs.toc)

HUMAN VISUAL SENSITIVITY GUIDED BIT EMBEDDING FOR VIDEO STEGANOGRAPHY

[1]S. Balu, [2]K. Amudha and [3]D.C. Nelson Kennedy Babu

[1]Department of IT, K.S. Rangasamy College of Technology, Tiruchengode, India
[2]Department of ECE, K.S. Rangasamy College of Technology, Tiruchengode, India
[3]CMS College of Engineering, Namakkal, Tamil Nadu, India

Received 2013-05-03, Revised 2013-06-19; Accepted 2013-06-24

ABSTRACT

While watching a video human visual system gives more attention on the foreground objects than background objects. That is, human vision system pays more attention to the region of interest, such as the human faces in the video content. Most of the video steganography algorithms embed secret information in video by considering every part of the video frames with equal importance. So the capacity of the steganography could not be increased to maintain low visual distortion. The proposed system will detect the foreground and it will allocate different bit rates for different regions. By doing this the visual distortion will be maintained and the capacity can be increased up to 33%.

Keywords: Steganography, Video, Bit Embedding, Security, Capacity

1. INTRODUCTION

Steganography is an art of hiding secret information in transmission components like text, image, audio, video and animation (Saravanan and Neeraja, 2013; Saravanan *et al*., 2013; Solanki *et al*., 2006). Generally, all the transmission components are stored in the storage devices as binary values. The binary values can be altered to cover secret information. Altering few bits may not amend originality of the transmission components in wide quantity, however if the changes are too high then, originality of the transmission components will get spoil. Thus to cover few kilobytes of data we need few mega bytes of transmission components. Combining Steganography and cryptography can improve the security dramatically.

Generally video Steganography can enter secret information into the video content by considering each part of the video frames with equal importance. Therefore the capacity of the steganography can be restricted to keep up low visual distortion.

The media with secret information is named as stego media and without hidden information is named as cover media (Ma and Zhang, 2002; Battiato *et al*., 2007; 2008; Chen and Ngan, 2007). Steganalysis is a method of extracting information from the stego media. Steganalysis is simply opposite to Steganography.

The volume of digital video has fully grown enormously in recent years and has become a serious information storage and exchange media. Transmission plays an awfully necessary role in computing and communication environments, with varied applications in recreation, advertising, distance learning, tourism, distributed CAD/CAM, GIS, sports (Abburu, 2010). Any video is formed from frames and every frame is formed from pixels. Every constituent represents a color value and depends upon the video the pixel sizes may vary from one bit to four bytes. These pixels are to be stored in computer hardware in binary type.

Let us think about a five second video with twenty five frames per second and every frame contains eighty thousand pixels (400×200) with pixel size 8 bits for discussion. The constituent with eight bit size is used to represent 256 completely different colors, vary from 00000000 to 11111111. The entire range of bits within the video content (That is Size) will be calculated as:

$$\text{Size} = \text{VL X FRX FS X BR}$$

Corresponding Author: S. Balu, Department of IT, K.S. Rangasamy College of Technology, Tiruchengode, India

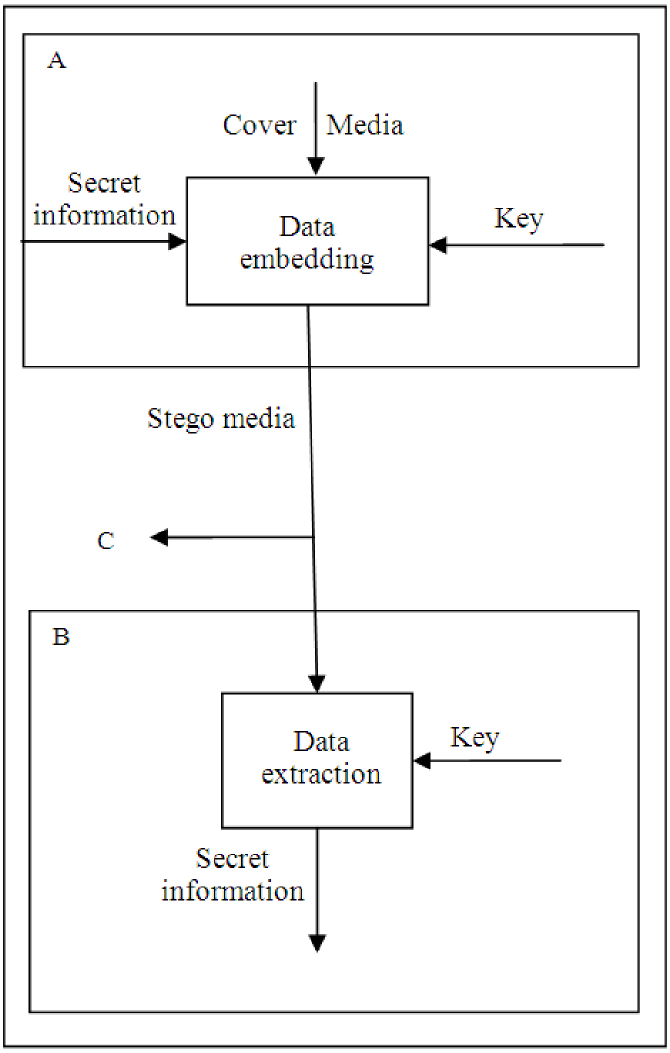

Fig. 1. Basic steganography

where, VL is that the video length, FR is that the frame rate, FS is that the frame size and BR is bit rate. For the on top of same video, Length is five Second, frame Rate is twenty five, frame size is eighty thousand and bit rate is 8. So, Size of the video will be calculated as:

$$\text{Size} = 5 \times 25 \times 80000 \times 8 = 80000000 \text{ Bits}$$

Unremarkably for human eyes the color 11110000* and 11110001* can seem like similar, since the distinction is just too low. Which suggests that the amendment in least vital bit (*LSB) might not be noticed by human eyes. If we tend to alter #11110000 as #01110000, then the color of a pixel can amendment to a different color. That's the amendment within the Most Significant Bit (#MSB), can amendment the color dramatically. This will be simply notified by everybody and also the originality of the image is going to be spoiled. Thus it's clear that the key information should be kept within the LSBs and not in MSBs of the cover media to cut back the detectable distortion.

Let us think about a scenario (Saravanan and Neeraja, 2013) that someone "A" needs to send secret information to a different person "B" and also the secret information is "Bring Files on Sunday". This information mustn't be identified by "C" who is knowledgeable in hacking. During this scenario A should take a video of size eight times bigger than secret information as shown in **Fig. 1**. He should convert the key information into binary type then he should store the key information within the LSBs of every constituent one by one. The resultant video (stego video) is going to be sending to B via network. Currently C can catch the packets using varied hacking tools accessible within the market and he will construct the video send by A to B (This is feasible as said in chapter 2). Currently C will simply see the video and he may think that there's no secret in their communication. This is how one can cheat the hackers using Steganography.

The performance of a Steganography is measured by 3 factors. They're security, capacity and Detectable Distortion (DD). The security should be high, so the active attacks and passive attacks mustn't reach finding secret information. Security will be achieved by ever-changing approach of storing the bits, for instance rather than storing first bit, second bit and third bit of secret information into LSB of 1st, second and third pixels, it will be keep in second, fourth and sixth pixels. By doing such variation within the bit alteration might confuse the steganalyst.

The capacity of the steganography will be calculated as:

$$\text{Capacity} = \left(\text{VS} / \text{BR}\right)\ \text{X NL}$$

where, VS is that the Video Size, BR is that the bit rate and NL is that the range of LSBs altered to represent the key information (This will be from one to three, for low distortion one will be used).

So, capacity of the on top of same video will be calculated for 1LSB alteration as:

$$\text{Capacity} = \left(80000000 / 8\right)\ \text{X } 1 = 10000000 \text{ Bits}$$

Similarly, for 2LSBs alteration the capacity will be calculated as:

$$\text{Capacity} = \left(80000000 / 8\right)\ \text{X } 2 = 20000000 \text{ Bits}$$

Normally capacity of the Steganography (Saravanan and Neeraja, 2013) will be enhanced by sterilizing additional LSBs of a constituent. For instance rather than ever-changing just 1 LSB of a constituent, if we tend to alter 2 LSBs then the capacity can become double. However increasing the capacity on the far side bound level can produce detectable distortion within the stego video. While maintain the low detectable distortion the capacity should be relaxed. The DD of a Steganography should be low, that is, the stego video mustn't have high visual artifacts (That is that the amendment within the video content mustn't be notified simply by the humans).

In this study, we tend our focusing on capacity and DD of the steganography method. The flexibility for human vision system to notice alterations in video sequences should be thought-about for correct bits embedding throughout steganography method. The capacity and DD will be controlled by characteristic the regions with varied distortion levels and by embedding additional bits within the unwanted regions.

This study is organized as follows. In section 2, the security issues in computer networks are analyzed and also the need for security mechanism is identified. In section 3, the foreground objects within the videos are identified by the motion attention index. The variation ranges are calculated in section 4, the proposed method of allocating bits is delineate in section 3. Section 6 analyzes the proposed scheme. Finally conclusions are given in section 7.

2. SECURITY ISSUES IN COMPUTER NETWORKS

As discussed by (Saravanan and Neeraja, 2013), nowadays everybody use computer networks and Internet to share resources and to exchange data between the connected nodes. This network will be classified into many sorts by supporting the properties like protocol, topology and design. Primarily based upon the required security level, price issue, performance and implementation limits anyone of this sort will be used. Topology defines the physical arrangements of the nodes of a network. Wide better-known topologies are bus, ring, star, mesh and hybrid. In topology all the nodes can connected employing a single cable (this cable can act as a backbone). Damaging the cable can cause network failure. The information will be simply hacked by hackers by recording the cable anyplace within the network. This is often a straightforward and low cost topology to implement. In ring topology the nodes of a network can connect via a hoop like cable and it is good in speed and information can be hacked easily by taping the cable, comparing bus topology.

In star, the network devices like hub or switch are accustomed by connecting all the nodes of a network. Taping of single cable might not be helpful to hack all the information of all the nodes if the network uses switches, as a result of switch merely forward the frame to a particular port, which is connected to specific node of the network. Just in case if a network uses hub, then recording one cable is enough to watch or hack the information of a network. Several hacking, network observance and packet capturing tools are obtainable within the market. This may break the protection in both wired and wireless networks. Mesh network topology connects all the nodes of a network to every different. As a result, it desires a lot of variety of cables and network adapters. Here the advantage is failure of single cable might not have an effect on the network performance and also the networks are highly stable. Throughout the information exchange the data can travel in multiple path, thus hacking is hard than previous topologies.

Using architectures we are able to classify the networks. Wide better-known architectures are peer-to peer and client-server. In peer-to-peer all nodes will communicate with one another without any specific server node to regulate. It's appropriate for little firms or establishments wherever the quantity of nodes is in smaller amount. Usually this sort of design accustomed to share resources like storage capacity, internet, printers, scanners and different things. Most of the little firms and DTP centers use this sort of network for simple installation. In client-server design a particular node will act as a Domain Controller that controls all the nodes of a network. All the nodes will login to the server to urge a particular service. Through the login method the nodes should send the user name, password and different data like text by captcha for correct authentication. The client-server design is nice just in case of security comparing peer-to-peer. Due to security reasons several massive firms uses this sort of design. Also, the protection risk is extremely high in wireless networks than wired networks, since the signal spreads over air (Saravanan and Sumathi, 2012a; 2012b), the hackers will sit anyplace within the coverage space to hack the information. So, it's clear that providing security to the information may be a terribly massive difficult task in computer networks and also the need for developing a security mechanism is must.

3. MOTION ATTENTION INDEX

Normally human attention on an object is directed by brain to complete a task. But in some situation the object can attract the humans and make them to listen, statically or dynamically. The static attraction is suitable for images and the dynamic attraction is more suitable for videos. So we can use the dynamic model discussed in (Ma and Zhang, 2002). This is a simple model which is used to detect the moving object in a video, with considering the overall motion. This model uses three different values namely intensity inductor (Used to detect moving object in a video without considering global motion), spatial coherence inductor and temporal coherence inductor (Considers camera motion). The Intensity Inductor Value (IIV) for a macroblock (i,j) of nth frame is calculated as:

$$IIV_{nij} = \frac{\sqrt{mvx^2_{nij} + mvy^2_{nij}}}{maxl_n}$$

where, $maxi_n$ is the maximum motion vector intensity in nth frame, $mvx^2{}_{nij}$ and $mvy^2{}_{nij}$ are the motion vectors in the nth frame. This IIV_{nij} is not sufficient because the camera movement can cause large intensities that cannot be calculated by this. To avoid such negative effect we must go for spatial inductor and temporal coherence inductor. The Spatial Coherence Index Value (SCIV) can be calculated for a macroblock (i,j) of nth frame as:

$$SCIV_{nij} = -\sum_{b=1}^{n_s} pd_n(b)Log(pd_n(b))$$

where, $pd_n(b)$ is a probability function, n_s is the number of histogram bin. Similarly the temporal Coherence Inductor Value (TCIV) can be calculated by:

$$TCIV_{nij} = -\sum_{b=1}^{n_s} pd_n(b)Log(pd_n(b))$$

where, $pd_n(b)$ is a probability distribution function and nt is the number of bin for motion direction histogram. Now we have both the values which is required to compute the Motion Attention Index Value (MAIV) for the macroblock (i,j). The MAIV values are used for detecting foreground and background in a video.

$$MAIV_{nij} = IIV_{nij} X\ TCIV_{nij} X\ \left(1 - IIV_{nij} X\ SCIV_{nij}\right)$$

The $MAIV_{nij}$ will be in between 0 and 1

4. REGIONS WITH VARIATIONS

In section 3, the detection of foreground and background in a video has been done. Based on this detection we can embed more number of bits by altering more number of LSBs in background than foreground. By doing this the capacity of the steganography process will be improved without increasing VD. But, still more bits can be embedded and VD can be improved further by adding the idea proposed by (Saravanan and Neeraja, 2013), with this existing idea. The proposed idea is to examine the blocks in each frame of the video. Each block can be compared with neighbor pixels for the variations. If the variation is too high, then the block can be marked to embed more bits. Similarly if the variation is too low then the block can be marked to allocate fewer bits. This variation can be calculated in such a way that the maximum value should be 1 and minimum value should be 0.

The pixels in the macroblock (i,j) will be compared with each other as follows to calculate the variations (Saravanan and Neeraja, 2013):

$$VR_{nij} = VR1_{nij} + VR2_{nij}$$

where, VR_{nij} is the variation range value, $VR1_{nij}$ and $VR2_{nij}$ will be calculated as follows:

$$VR1_{nij} = \sum_{y=1}^{maxy} \sum_{x=1}^{maxx-1} \sum_{x1=x+1}^{maxx} (P_{nij}(x,y) - P_{nij}(x1,y))$$

$$VR2_{nij} = \sum_{x=1}^{maxx} \sum_{y=1}^{maxy-1} \sum_{y1=y+1}^{maxy} (P_{nij}(x,y) - P_{nij}(x,y1))$$

where, maxx is the maximum number of pixels in x-axis and maxy is the maximum number of pixels in y-axis of a macroblock respectively. $P_{nij}(x,y)$ is the pixel value at location (x,y). Using the VR_{nij} value the smoothness or randomness of the macroblock will be detected. For smooth macroblock the VR_{nij} value will be less and for highly variation macroblock the VR_{nij} value will be large (Saravanan and Neeraja, 2013). This range of values can be converted to VR'_{nij}, the range between 0 and 1.

5. PROPOSED SCHEME

To develop a good steganography algorithm the ability of human eyes to detect the variations in the video should be considered. The basic idea of the proposed scheme is to embed less number of bits for foreground objects and to embed more number of bits for background objects. The video can be split into number of macroblocks and then we have to find whether the macroblock contains foreground object or a background object by motion attention index and variation range. The motion attention index can be calculated by section 3 and variation range can be calculated using section 4. The Human vision system's Region of Interest (HROI) value can be calculated by using both motion attention index and Variation Range (VR) values. Then based on this value the steganography algorithm can take a decision on, how much bits of secret information to be embedded for a particular block. The HROI value will be calculated as:

$$I_{nij} = \begin{cases} MAIV_{max}, & \text{if} \quad MAIV > \tau \\ VR'_{nij} & \text{if} \quad VR'_{nij} > t \\ FV & \text{otherwise} \end{cases}$$

where, τ and t are thresholds for indicating visual attended regions (value between 0 and 1), FV is the fixed value. $MAIV_{nij}$ is motion attention index of a macroblock at location (i,j) and VR'_{nij} is variation range of a macroblock at location (i.j).

For each macroblock the HROI will be calculated using the above equations. These values of HROI will be given as an input for steganography algorithm, so that it will decide the amount of LSB's to be altered for embedding secret information in a particular macroblock. If the HROI is too high then less number of LSBs should be altered for hiding secret information, else if the HROI is too low then more number of LSBs should be altered to hide secret information. The decided number of LSBs will be always less than the maximum number of LSBs allowed or it will be equal, never greater than the maximum number of LSBs allowed. Hence the VD will get reduced dramatically in the stego video.

6. PERFORMANCE ANALYSIS

In this section, we have analyzed the capacity, DD and Security improvement between the existing fixed LSBs method with our proposed method. Let us consider a video of length 15 min with the frame rate of about 29 frames per second, frame size of about 84480 pixels (352×240 = 84480), bit rate of about 8 bit per pixel for discussion. The video size can be calculates as:

$$\text{Size} = 15 \times 29 \times 84480 \times 8 = 97996800 \text{ Bits}$$

The capacity of the 1LSB method can be calculated as:

$$\text{Capacity } 1 = (97996800 / 8) \text{ X } 1 = 12249600 \text{ Bits}$$

The capacity of the 2LSB method can be calculated as:

$$\text{Capacity } 2 = (97996800 / 8) \text{ X } 2 = 24499200 \text{ Bits}$$

Similarly, the capacity of the 3LSB method can be calculated as:

$$\text{Capacity } 3 = \left(97996800 / 8\right) \text{ X } 3 = 2351923200 \text{ Bits}$$

From the above values we can say that Capacity 1 < Capacity 2 < Capacity 3 is true. That is by increasing the number of LSBs to hide the data will increase the capacity. Also, we know that by increasing the capacity the DD will also get increased. That is DD of 3 LSBs method will be greater than the DD of the 2 LSBs method and 2 LSBs method will be greater than DD of 1 LSB method.

So, we can say:

$$\text{DD3} > \text{DD2} > \text{DD1 is true}$$

where, DD3, DD2 and DD1 represent the detectable distortion level of 3LSB, 2LSB and 1LSB methods respectively.

Our proposed scheme uses different number of LSBs for different macroblocks based on HROI value as discussed in section V. Different number of LSBs in different regions of the image makes better security, since the information like block size, total number of blocks and number of LSBs to be read from a specific block are unknown to the hacker. Based on HROI value for a block the either 1 or 2 or 3 LSBs can be altered. For the blocks that contain foreground objects 1 LSB method can be applied and for the blocks that contain background object 3 LSBs method can be applied. If any block contain background object and plain region then 2 LSBs method can be applied. Hence the capacity of our proposed method will be less than or equal to the 3 LSBs method. Also, we can say that the capacity of our proposed method is greater than or equal to the 1 LSB method.

So, the below condition is true:

$$\text{CL1} \le \text{CPM} \le \text{CL3}$$

where, CL1 and CL3 is the capacity of 3 LSBs method and 1 LSB method, CPM is the capacity of our proposed method. In worst case (If the entire video has foreground object) the CPM = CL1 is true and in best case (If the entire video has background object and contains no plain region) the CPM = CL3 is also true.

Similarly, the DD of our proposed method can be represented as:

$$\text{DD3} \le \text{PDD} \le \text{DD1}$$

where, DD3 and DD1 is the detectable distortion level of 3 LSBs and 1 LSB methods respectively, PDD is the detectable distortion level of our proposed method. Generally, no video comes with 100% foreground objects. Probably the amount of region attended by human vision system will be lesser than the region not attended by the human vision system. Hence, our proposed algorithm will give optimal result in capacity and excellent result in DD comparing existing methods perceptually.

7. CONCLUSION

In this study we have proposed a VDSI based steganography. This scheme makes the steganography process to embed less number of bits for foreground, where the viewers can easily identify the distortion and more number of bits for background, where the viewers cannot easily identify the distortion. Thus the capacity gets improved without affecting the perceptual quality. That is without affecting the VD. In future, the face detection algorithms can be added along with this to embed less number of bits where the human faces are available in frames. This will further improve the VD.

8. REFERENCES

Abburu, S., 2010. Semantic segmentation and event detection in sports video using rule based approach. Int. J. Comput. Sci. Netw. Security, 10: 40-45.

Battiato, S., C. Guarnera, G.D. Blasi, G. Gallo and G. Puglisi *et al*., 2008. A novel artificial mosaic generation technique driven by local gradient analysis. Proceedings of the 8th International Conference on Computational Science, Jun. 23-25, Springer-Verlag Berlin, pp: 76-85. DOI:10.1007/978-3-540-69387-1_9

Battiato, S., G.D. Blasi, G.M. Farinella and G. Gallo, 2007. Digital mosaic frameworks-An overview. Comput. Graph. Forum, 26: 794-812. DOI: 10.1111/j.1467-8659.2007.01021.x

Chen, Z. and K.N. Ngan, 2007. Towards rate-distortion tradeoff in real-time color video coding. IEEE Trans. Circ. Syst. Video Technol., 17: 158-167. DOI: 10.1109/TCSVT.2006.888022

Ma, F. and H.J. Zhang, 2002. A model of motion attention for video skimming. Proceedings of the International Conference on Image Processing, Sept. 22-25, IEEE Xplore Press, pp: 129-132. DOI: 10.1109/ICIP.2002.1037976

Saravanan, V. and A. Neeraja, 2013. Security issues in computer networks and stegnography. Proceedings of the 7th International Conference on Intelligent Systems and Control, Jan. 4-5, IEEE Xplore Press, Coimbatore, Tamil Nadu, India, pp: 363-366. DOI: 10.1109/ISCO.2013.6481180

Saravanan, V. and A. Sumathi, 2012b. Handoff mobiles with low latency in heterogeneous networks for seamless mobility: A survey and future directions. Eur. J. Sci. Res., 81: 417-424.

Saravanan, V. and D.A. Sumathi, 2012a. Dynamic handoff decision based on current traffic level and neighbor information in wireless data networks. Proceedings of the IEEE International Conference on Advanced Computing, Dec. 13-15, IEEE Xplore Press, Chennai, pp: 1-5. DOI: 10.1109/ICoAC.2012.6416797

Saravanan, V., A. Sumathi, S. Shanthana and M. Rizvana, 2013b. Dual mode mpeg steganography scheme for mobile and fixed devices. Int. J. Eng. Res. Dev., 6: 23-27.

Solanki, K., U. Madhow, B.S. Manjunath, S. Chandrasekaran and I. El-Khalil, 2006. ‘Print and Scan’ resilient data hiding in images. IEEE Trans. Inform. Forens. Security, 1: 464-478. DOI: 10.1109/TIFS.2006.885032