A modified collage steganography using modified fuzzy C-means algorithm on images

Dr. R. Udayakumar,
Associate Professor, Department Of IT, Bharath University

Abstract-- we propose a modified collage steganography using the modified fuzzy c-means algorithm on images. It consists of object images and cover image. Object image consists of the secret message which is to be embedded on the cover image. A proper cover image has to be chosen based on the prior analysis. The fuzzy based approach has been formulated for each object image which is to be placed on cover image. The cover image should possess the high probability of embedding the object images. The cover image has to be chosen such that it has more smooth regions and should possess a high degree of relationship on the object images which is to be embedded. The smooth regions are clustered by the fuzzy c-means algorithm which has a greater capability of clustering. The cover image is transformed by placing the object images. The information is retrieved using the different template matching techniques.

Keywords- Information Security; Text Steganography; Information Hiding; Text watermarking; Linguistic Text Steganography

I. INTRODUCTION

As the number of data being exchanged on the Internet increases, the network security is becoming more and more important. Therefore, the confidentiality and data integrity are required to protect against unauthorized access. This has resulted in an explosive growth of the field of information hiding. Various methods like Cryptography, Steganography, coding and soon have been used for this purpose. However, during recent years, Steganography perhaps has attracted more attention than others.

Steganography is the method of hiding information such that its presence cannot be detected. A secret message is encoded in such a manner that the existence of the information is hidden. Combining with existing communication methods, Steganography can be used to carry out hidden exchanges data. It is not to keep others from knowing the hidden information, but it is to keep others from suspecting that the information even exists. A Steganography method fails if it causes a suspicion that there is secret information in a carrier medium [4].

The objective of Steganography is to establish a secured communication in a completely undetectable manner and to avoid drawing suspicion to the transmission of a hidden Image segmentation is the process and technology that to divide images into the regions with distinctive features and extract the target interested in. Features can be gray, color, texture and so on, target can correspond to single region, may also correspond to a number of regions.

Unsupervised clustering method used regularly for the fuzzy boundary region partition, like K-Means, Fuzzy C-Means [4-8], ISODATA, etc. FCM is more effective to the fuzzy boundary region segment, but the biggest disadvantage is that no better way to determine the C value of clustering and the initial cluster centers, essentially, FCM is a local search optimization algorithm, it will converge to the local minimum point and this clustering effect would have a greater impact if the initial selection value are not properly [12, 13].

In this paper, the new method of the color image segmentation was put forward based FCM. Firstly, using evolution clustering method (ECM) for the clustering partition according as the RGB value of image pixel, the gained centers as the initial cluster centers of parallel FCM, using parallel FCM to optimize the cluster centers to complete the partition of fuzzy clustering, through the elimination of fuzziness into a certainties classification, realizing the clustering segmentation to meet the elements of various cluster in a high similarity. Large amount general images and remote sensing images segmentation have verified this method, the manual evaluation result showed that the accuracy is higher and improve the remote sensing image segmentation efficiency.

II. GENERALIZED COLLAGE STEGANOGRAPHY

Generalized collage steganography is an extension of Collage steganography proposed in [16]. This method hides information by changing the appearance instead of the features of the image. The steganographic message needs to be converted to binary data prior to embedding. In order to begin the steganographic process, two sets of image databases are prepared, i.e., the cover image database and the object image database. The former collects the images which are potential carriers, while the latter includes directories of transparent object images which are candidates to be pasted on top of the cover. The object images are of file type PNG or GIF and the sizes are within a set of constraints when establishing the image database. The object image file types containing transparent properties have irregular shape. This allows the object image files to be integrated into other image file types without the visible
rectangular borders. The image sizes in both databases are not limited. However, the transparent object images require an art as well as a science when used in conjunction with the cover images.

A. Image Coordinate Representation

Prior to presenting the generalized collage steganographic method, the image coordinate representation is described as follows. Suppose an image is of the size $M \times N$, which indicates the image has $M$ rows and $N$ columns. The image pixel coordinate is defined as index $(x,y)$, where $x$ indicates the row index and $y$ indicates the column index. The upper left pixel is located at $(1,1)$ and index $(M,N)$ is the bottom right pixel.

A pixel, say $p$, can be represented by a vector using homogeneous coordinates, i.e. $p = [x, y, 1]^T$. Such coordinate system extends one dimension from the original system space. This allows simpler representation when working on the affine transformation which will be used in this paper.

B. Affine Transformation

The affine transformation is a two dimensional geometric transformation representing, in general, a set of mapping operations. This subsection discusses three operations, scaling, translation and rotation, which will be used in this paper. A pixel index represented as a vector, $p_0 = [x_0, y_0, 1]^T$, the pixel can be transformed to another location $p_1$ in the coordinate system using the affine transformation matrix $T$. As shown in (1), $p_1$ can be obtained by matrix multiplication.

$$p_1 = T \cdot p_0$$

where $S_x$ and $S_y$ are the scaling factors, $t_x$ and $t_y$ are the translation factors, zooming and shifting along rows and columns respectively, and $\theta$ is a rotation of radians counterclockwise about the origin. If $\theta = 0$ and $S_x = S_y = 1$, a shape merely operates translation movement from one location to another in the coordinate space. The pixel values after the transformation, especially scaling and rotation, is calculated using the inverse mapping approach described in.

The Smooth regions of the reference are formulated using the FCM which is a clustering algorithm.

III. FUZZY C-MEANS ALGORITHM

Fuzzy C means is a method of clustering which allows one piece of data to belong to two or more clusters. It is based on the optimization of objective function. By iteratively updating the cluster centers and membership grades for each data point, Fuzzy C means moves the cluster centers to the right location within the dataset.

Fuzzy C-means is a clustering algorithm that used membership degree to determine each data point belongs to a certain cluster. FCM divided the $n$ vectors $x_i, i = 1, 2, ..., n$ into $c$ fuzzy group, and computing the cluster center of each group, making value function of non-similarity index to achieve the minimum. FCM algorithm making each of the given data points with values between 0,1 membership to determine its degree of belonging to various groups through fuzzy partition. And to suit the introduction of fuzzy partition, the membership matrix $U$ allowed the element value between 0,1.

In addition to normalizes the membership degree sum of a data set equivalent to $1$:

$$\sum_{i=1}^{n} u_{ij} = 1, \forall j = 1, ..., n$$

Then, the values function of FCM (objective function) as follows:

$$J(U, c_1, ..., c_c) = \sum_{i=1}^{c} \sum_{j=1}^{n} \left( \sum_{i=1}^{c} u_{ij} \right)^{m} (d_{ij})^{2}$$

$u_{ij}$ ranged between 0,1 here, $c$ is the cluster center of fuzzy group $i$, $d_{ij} = |x_i - c_i|$ is the Euclidean distance between the first $i$ cluster center with the first $j$ data point. $m \in [1, \infty)$ is a weighted index.

The new constructed objective function can be obtained necessary condition so that Eq (2) to achieve the minimum:

$$J(U, c_1, ..., C_c, \lambda_1, ..., \lambda_c) = J(U, c_1, ..., C_c) + \sum_{j=1}^{n} \lambda_j \left( \sum_{i=1}^{c} u_{ij} - 1 \right)$$

In Eq (3), the $\lambda_j (j = 1, 2, ..., n)$ is Lagrange multiplier of $n$ of constrained formulas of Eq (1). Derivation of each input parameter, so that Eq (2) to reach the minimum necessary conditions are as follows:

$$c_i = \frac{\sum_{j=1}^{n} (U_{ij})^m \cdot x_j}{\sum_{j=1}^{n} (U_{ij})^m}$$

and

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left( \frac{d_{ij}}{d_{kj}} \right)^{2(m-1)}}$$
From the above two necessary conditions, Fuzzy C-Means is a simple iterative process. In the run-time of batch mode, FCM with the following steps to determine the cluster centers c and the membership matrix U:

**Step 1.** Using a random number in value between 0, 1 to initialize the membership matrix U, to meet constraints of Eq (1).

**Step 2.** Computing the cluster centers c, (i=1, 2, ..., c) by using Eq (4).

**Step 3.** Calculating the values function with Eq (2) The algorithm to stop if it is less than a determined threshold or its previous value of the relative change in function value is less than a certain threshold.

**Step 4.** Calculation of the new matrix U by Eq (5), Return step 2.

The above algorithm can also initialize cluster center firstly, and run the iterative process after. The algorithm does not guarantee an optimal solution to the convergence, so its performance depends on the initial cluster centers. Therefore, another fast algorithm can be used to determine the initial cluster centers, or start the algorithm with different initial clusters center every time, running FCM repeatedly.

When the algorithm convergence, it can get each clustering centers and the membership degrees of various clusters samples, completing the partition of fuzzy clustering, and finally through the elimination of fuzziness into a certainty classification, to achieve final clustering segmentation.

### III. PARALLEL FUZZY C-MEANS ALGORITHM

Improving the performance of the FCM based image segmentation by distributing computation and main memory usage was the main objective for design and implementation of this parallel FCM application.

This parallel FCM algorithm divides the image pixels equally among the processors so that each processor handles n/p data points (n is the total number of pixels and p is the number of processors involved in the computation). The fuzzy membership function (5) is distributed among the processors and is used for the calculation of the degree of membership, \( u_{ij}(x_i) \), only for the local data set. By dividing the data set among p processors the likelihood of carrying out the computation only on the main memory of the processors and without the need to access secondary storage will be increased. This enhances the performance and efficiency as compared to FCM algorithm.

The algorithm will take the following steps:

**On the initiating processor:**

1. Divide X into n/p subsets and send each subset to a participating processor using a one to all personalized broadcast.

2. One to all broadcast Vector Y

On each participating processor:

3. Calculate:

\[
A[j,i] = u_{ij}(x_i) = \frac{1}{\sum_{j=1}^{k} \frac{\| x_i - v_j \|^2}{\| v_{i,j} \|^2}}
\]

Where 0 ≤ i ≤ k – 1, \( (n/p) * P_{id} \) ≤ i ≤ \( (n/p) * P_{id} + 1 \), and 0 < P_{id} < p - 1. \( A[j,i] \) is a member of matrix A, partial membership matrix.

4. Calculate:

\[
Q[l] = \frac{n/p}{\sum_{J=1}^{n/p} (A[j,l])^m}
\]

And form vector Q from Q[l] for 0 < L < n/p, Vector Q would be partial denominator of the equation that calculates V in step 5.

5. Calculate

\[
L = A^m \ast \left[ \begin{array}{c} X_n P_{id}/p \\ (n P_{id}/p + 1) \\ \vdots \\ X_n (P_{id} + 1)/p \end{array} \right]
\]

Where L is a vector of K values form the partial numerator of the equation that calculates V in above step.

6. Perform all reduce operation for L and Q to form final numerator \( L^* \)

7. For 0 ≤ I ≤ k-1 calculate:

\[
V[i] = L^* [i] \\
Q^* [i]
\]

If \( V(\text{current}) - V(\text{previous}) > E \) then go to step 2. If the difference between previous and new clusters’ center vector is less than E then stop.

To divide the image file among p processors, it is cut horizontally into equal blocks of h/p rows where h presents number of the rows in the image. The image height is not always an integral multiple of the number of processors. The remaining rows cannot be greater than n - 1,
therefore they are divided among the processors starting with the processor having the lowest ID. After the completion of the computation of the centers of the clusters and final degree of membership matrix, according to the parallel algorithm discussed above, each processor defuzzifies its local data and sends it back to the initiating processor. The initiating processor, in turn, writes the received data into the output file to form the segmented image.

Embedding and retrieving Methods
The secret message is embedded into several steg images which are found in the database. The appropriate steg images have to be chosen for the appropriate reference images. Initially the reference images are clustered using the FCM and the smoother regions are evaluated. The steg images are placed in the smoother regions appropriately after some affine transformation.

While retrieving the steg images the template matching techniques are used to retrieve the placed steg images and after the inverse affine transformation, the secret message is decrypted from the several steg images and arranged to get a meaningful secret message.

Note that if the location the object image is out of the cover image boundary, the selected object image can be resized. One of the methods is based on the ratio of the lengths between the object image borders and the cover image borders. However, the acquired scaling factor may be misled by such manipulation. Hence, the selection of the permissible area and the possibilities of transformation factors need to be considered. For retrieval, the message can be extracted by using template matching techniques described as in the previous section with the steganographic image and the information in the key file along with the quotient information bits hidden in the cover image.

(a)                                      (b)

CONCLUSIONS
The proposed algorithm for the image segmentation includes the combination of the ECM and parallel FCM. The initial cluster problem is solved which increases the performance of the FCM and to optimize clustering process. The parallel FCM decreases the time consumed by the FCM and the membership value is computed parallel.

REFERENCES