ABSTRACT

Steganography is the art of secret communication where secret information is embedded into a cover digital media. The main characteristic of steganography is that the very existence of the secret information within the cover media is not noticeable. Due to this, steganography has been used by many criminals for secret communication. Among the different media used for steganography, image has been a good choice since the minor variations in image parameters are imperceptible to the normal human eye. Hacking this secret communication is steganalysis, i.e. identifying a suspicious image as clear (cover) image or as stego image (one with secret data). This is a two class classification problem. Image steganalysis can be embedding specific or blind (universal) steganalysis. When steganalysis is done with the knowledge of the embedding algorithm, it is called embedding specific steganalysis. If steganalysis is done without the knowledge of the embedding algorithm, it is called universal or blind steganalysis.

The objective of this research work is to extract all possible changes in image features that have changed due to steganographic embedding and identify the best features by proper optimization techniques. Optimization with statistical and Bio inspired methods are implemented and classified with single classifiers (SVM and MLP) and fusion classifiers based on Bayes, Dempster Schafer and Decision template fusion schemes. Thus the aim is to DESIGN AN UNIVERSAL IMAGE STEGANALYSER
The existing research work, in this field shows that image steganography and steganalysis can be in spatial and transform domains. Since there are specific steganographic embedding methods in spatial and transform domain, feature extraction methods also differ in each domain. In both the domains, the classifier accuracy depends on the features selected. Extracting few features and optimizing them may lead to loss of vital embedding information resulting in low classification accuracy (steganalysis). To overcome this, recent researchers choose all possible variations in image features creating a large (rich) set of features. Though the large feature set captures all possible embedding artefacts, they suffer from dimensionality problem in terms of time and space. Another disadvantage of large dimensional feature set is overtraining of the classifiers that lead to poor classification (steganalysis) accuracy. Existing research works concentrate on feature extraction and optimization based on statistical methods. While these methods have less classification accuracies, bio inspired algorithms seem more promising. Bio inspired algorithms start the optimization with some initial candidate solutions, then improve these solutions iteratively till global minima (level of satisfaction). This research work has implemented Ant Lion Optimization (ALO) for optimizing the large dimensional image data. ALO is based on the hunting behaviour of antlions where the movements of the ants and antlions are random walks in search space and the fitness function depends on the size of the pits built by the antlions. Fitter ants are caught by the antlion.

The images used in this research are taken from the Break Our Steganographic System (BOSS) image database for implementation and testing. The most recent nsF5 steganographic algorithm has been used to create stego images in transform (JPEG) domain and S-UNIWARD is used for creating stego images in spatial domain.
Steganalysis in JPEG domain is targeted with four types of image features extracted from clean and stego images. The first feature set is 548 features of Calibrated Markov and histogram features of Discrete Cosine Transform (DCT) coefficients (CMDCT features), the second set comprises 8726 features of Co-occurrence Features of Difference in DCT Coefficients (COD features), the third set is 48600 high dimensional features of co-occurrences of selected DCT coefficients (DCT-300p features) and the fourth set is 23230 features comprising all combinations of co-occurrences of DCT coefficients (DCT23230 features). These features have been optimized by unsupervised cluster based optimization (statistical method) and Ant Lion Optimization (Bio Inspired optimization) separately. The optimized features are classified with single and fusion classifiers.

The unsupervised clustering optimization of COD features gives an average classification accuracy of 58.74% with Bayes classifier. The Ant Lion Optimization (ALO) of DCT-300p features give an average classification accuracy of 60.66% with Bayes fusion classifier, DCT23230 features gives an average classification accuracy of 64.44%. Thus in all cases, for JPEG steganalysis Bayes fusion classifier gives high average classification accuracy. In the spatial domain, 34671 high dimensional features are extracted from noise residuals of image pixels and optimized with unsupervised cluster based optimization and ALO optimization separately. While the classification accuracy for cluster method is 99.25%, for ALO method the classification accuracy is 99.3% for all fusion classifiers and SVM classifier. Thus ALO based optimization gives highest classification accuracies for all classifiers in spatial domain.
The results are evaluated against the recent research works in terms of classification accuracy, timing analysis and space complexity. Timing analysis indicates that the total running time of this research work is very less compared to existing researches in JPEG and spatial domains. Also, inspite of very high feature set reduction ratio (99% for DCT23230 features and 91% for S34671 features) the classification accuracy in each case is high. Based on these analyses, it can be concluded that ALO based Bayes fusion classifier acts as the best universal image steganalyser in both JPEG and spatial domains. Thus this research has implemented a Universal image Steganalyser (Bio-Steganalyzer) by computational intelligence.