Design of a Digital Forensics Image Mining System

R. Brown[1], B. Pham[1] and O. de Vel[2]

[1]Faculty of Information Technology,
Queensland University of Technology,
GPO Box 2434,
Brisbane 4001, Australia.
{b.pham,r.brown}@qut.edu.au

[2]Information Networks Division,
Defence Science and Technology Organisation,
PO Box 1500
Edinburgh 5111, Australia
Olivier.DeVel@dsto.defence.gov.au

Abstract Increasing amount of illicit image data transmitted via the internet has triggered the need to develop effective image mining systems for digital forensics purposes. This paper discusses the requirements of digital image forensics which underpin the design of our forensic image mining system. This system can be trained by a hierarchical Support Vector Machine (SVM) to detect objects and scenes which are made up of components under spatial or non-spatial constraints. Forensic investigators can communicate with the system via a grammar which allows object description for training, searching, querying and relevance feedback. In addition, we propose to use a Bayesian networks approach to deal with information uncertainties which are inherent in forensic work. These inference networks will be constructed to model probability interactions between beliefs, adapt to different users' retrieval patterns, and mimic human judgement of semantic content of image patches. An analysis of the performance of the first prototype of the system is also provided.

1 Introduction

Digital forensics is the application of computer analysis techniques to determine potential legal evidence of computer crimes or misuse that are caused by unauthorised users or by unauthorised activities generated by authorised users. The significance of digital forensics can be seen from the 2003 Computer Crime and Security Survey, published jointly by the Computer Security Institute and FBI, which reported total annual losses incurred by unauthorized computer use exceeding USD200 million for 251 organizations surveyed [1]. Digital forensics covers a wide range of applications such as law enforcement, fraud investigation, theft or destruction of intellectual property. Techniques used for such investigations are varied and may include data mining and analysis, timeline correlation, information hiding analysis, etc. Since multimedia format is widely used and readily available via the Internet, there are increasing criminal activities in the last few years, which involve the transmission and usage of inappropriate material such as child pornography in this format. Hence, much forensic evidence comes in the form of images or videos that contain objects and/or scenes that may be related to criminal behaviours. A typical investigation in digital forensics can generate large image and video data sets. For example, a disk can easily store several thousands of images and videos in normal files, browser cache files and unallocated space (i.e., non-file system areas on the disk which may contain fragments of files). It has been estimated that, as of late 2003, there exist some 260 million pages of pornography on the Internet [2]. This can make the task of searching for, and retrieving, images/videos very time consuming. Digital Image Forensics (DIF) efficiently seeks for evidence by using appropriate techniques based on image analysis, retrieval and mining. Owing to rising criminal activities via the internet, the use of such techniques for investigative purposes have only recently emerged, although they have been intensively researched over the last three decades for many other important applications: medical diagnosis, mineral exploration, environmental monitoring and planning, aerial surveillance, etc.

Content-based approaches have been developed that are based on some general low-level visual features such as colour, shape, texture e.g. [3]. Search-by-example is a common practice whereby an image is supplied and the system returns images that have features similar to those of the supplied image. The similarity of images is determined by the values of similarity measures that are specifically defined for each feature according to their physical meaning. Since the quality of the retrieval results relies on the choice of features and their similarity measures, much research has been focused on identifying features with strong discriminatory power and similarity measures that are meaningful and useful. In addition, we would ideally want a more "intelligent" system which can include high-level knowledge, deal with incomplete and/or uncertain information, and learn from previous experience. Such systems could include, for example [4]:

- Model-based Methods: A model of each object to be recognised is developed. These objects are classified using their constituent components that in turn are characterised in terms of their primitives,
- Statistical Modeling Methods: Statistical techniques are used to assign semantic classes to different regions/objects of an image, and
- User Relevance Feedback Methods: User feedback is required to drive and refine the retrieval process. The system is thus able to derive improved rules from the feedback and consequently generate better semantic classes of images.

Model-based methods exploit detailed knowledge about the object and are capable of reasoning about the nature of the object. However, the models created are often handcrafted and cannot easily improve their performance by learning. Statistical modelling techniques rely on statistical associations between image semantics and, as such, do not require the generation of any complex object model. Such associations can be learned using the statistical model. However, it is difficult for the investigator to interpret some of the results (e.g., "why are these objects in the image scene similar?") because statistical modelling techniques cannot easily reason with any high-level knowledge about the regions and image scene. User relevance feedback techniques inherently capture continuous learning as the system is able to build up a knowledge base of past user feedback. Quite elaborate feedback mechanisms can be implemented, e.g., ranking of images, input from collaborating investigators etc. (e.g., [5]). Image mining in digital forensics would ideally use a combination or hybridization of these methods.

In Section 2, we discuss various requirements of image forensics in terms of types of search, level of performance, learning ability and user interfaces. Section 3 presents the operation model of our forensic image mining system and the motivations behind its design. Section 4 gives an overview of how an SVM is used for training to detect objects and scenes which are described as a hierarchy of components and constraints, while Section 5 briefly describes the grammar which supports the modes of interaction between users and the system for specification, querying and relevance feedback for continual improvement. A summary of performance analysis of the prototype implemented so far is also provided. More details can be found in our two previous papers [6, 7].

2 Requirements of Forensic Image Mining

Image mining is only one of many different activities undertaken during a digital forensic investigation. As mentioned previously, a digital forensic investigation can involve a large number of data/evidence derived from a variety of sources as, for example: structured and unstructured files (e.g., text, marked-up text, databases), images, videos, music, network packets and router tables, process tables, telephone call records and so on. Also, an investigation may involve access to partial data (such as disk clusters), hidden data (e.g., data in disk partition gaps, steganography), encrypted data etc. The basic process in an investigation involving digital evidence would consist of a sequence of rigorous steps, including: extracting all of the data whilst maintaining the integrity of the original media and ensuring the chain of custody, filtering out the irrelevant data and identifying the useful data and metadata (e.g., file timestamps), deriving timelines, establishing the relationships between the disparate data (link analysis and link discovery), establishing causal relationships (causal analysis), identifying and extracting profiles, generating a comprehensive report etc.

The challenge in digital forensics is to find and discover forensically interesting, suspicious, or useful patterns or partial patterns in the potentially very large (now of the order of terabytes, TB) data sets. This task is analogous to the "needle-in-the-haystack" problem or, in the case of partial patterns located in multiple sources of evidence, "bits-of-needles-in-bits-of-haystacks". Furthermore, digital forensics has some unique requirements that make it rather different from traditional pattern extraction activities, for example [4]:

- Digital forensics deals with data instances that are both unrelated and related. That is, data instances may have multiple relations (e.g. networks of computer users, email cliques, geographical co-location etc.).
- The "interestingness" of data or sequences of events may be determined their low frequency of occurrence and possibly their non-repetitiveness. Unusual events may be more relevant in an investigation (i.e. we may be interested in the 'outliers').

- Sources of data in digital forensics are large, thereby requiring the consolidation of multiple data sources.
- Data sources may be high-dimensional and involve very different and sparse attributes.
- False negatives need to be minimised as the cost of "missing the needle in the haystack" is large. On the other hand, the number of false positives is not an overly sensitive parameter though, clearly, it should be kept to a minimum.

3 Operational Model

In order to design and implement an efficient image mining system architecture, an operational model of the digital forensic image mining process was developed. This model reflects the procedures undertaken by an investigator during a typical digital forensics investigation.

The model consists of two "activities", namely one involving the rapid reduction of the large quantity of evidence that is involved in a case, and one involving the core image mining activities that deal with the actual image retrieval process for digital forensic examination. The former activity, as mentioned in Section II, involves the execution of a chain (in reality, a forest of connected trees) of forensic tools for analyzing the content of large data streams (disks and other data), filtering the data streams for data reduction, extracting meta-data (eg, file timestamps) etc. to downstream analysis and decision making that leads to a successful investigation. The latter activity is simply one of the many possible forensic tools deployed in the case investigation graph. The core image mining operational model follows two stages, namely the training phase and the testing or classification phase.

The training phase, also referred to as the classification model-generation phase, builds the object models relevant to the particular domain at hand. This phase is usually undertaken by an experienced investigator who has an insight into the object types involved in the particular case under investigation, an understanding of the classifier, knowledge of the object layout (eg, constraints such as positions, orientations etc.) and so on. The investigator will also be responsible for providing the relevance feedback on *a priori* evidence (eg, images from similar cases) in order to refine and improve the quality of the classification model. We propose to use a Bayesian Network for query refinement with a set of relevance feedback parameters (see Section VI). The testing phase uses the refined classification model (given by the set of model parameters) developed during the training phase to classify the set of images found in the case under investigation.

We have designed and developed a complete operational system for digital forensics which implements both the digital forensic examination process (the chain of forensic tools) as well as a prototype model-building and classifier system that focuses on the core image mining component of the operational model. The digital forensics investigative system (called "CFIT", or computer forensics investigative toolkit) is not described here. In this paper, we focus on the model-building and classifier system.

4 Detection of Component-based Objects and Scenes

There have been various component-based systems which deal with human detection. For example, features such as eye, nose, and mouth are first detected and then combined in a spatially constraint configuration in order to determine a face e.g. [8]. Other systems detect humans and their actions for various purposes: surveillance (e.g. detection of criminal activities [9]; movement recognition (e.g. gesture recognition for interactive dance systems [10]. The underlying models for such methods can be grouped in two main categories: task-specific models and general models that can be applied to specific tasks. The task-specific approach constructs a model from the components of a human silhouette and tightly coupled it with constraints that govern a specific action of interest e.g. [10], [11], and [12]. This approach is rather restrictive and does not provide a framework that can be readily extended in order to model different behaviours for other applications. The general approach, on the other hand, constructs a model from primitives in a bottom-up fashion and uses a regular grammar to represent various modes of motion and interactions e.g. [13], [14], and [15]. The system is then trained using models that represent certain exemplar behaviours. A special type of statistical models called Hidden Markov Models (HMM) [16] is used to represent both *a priori* knowledge and new knowledge resulted from new behaviours. Low level primitives are firstly detected before they are passed into the grammar for behaviour analysis. These systems, although robust, rely on motion information to resolve ambiguities.

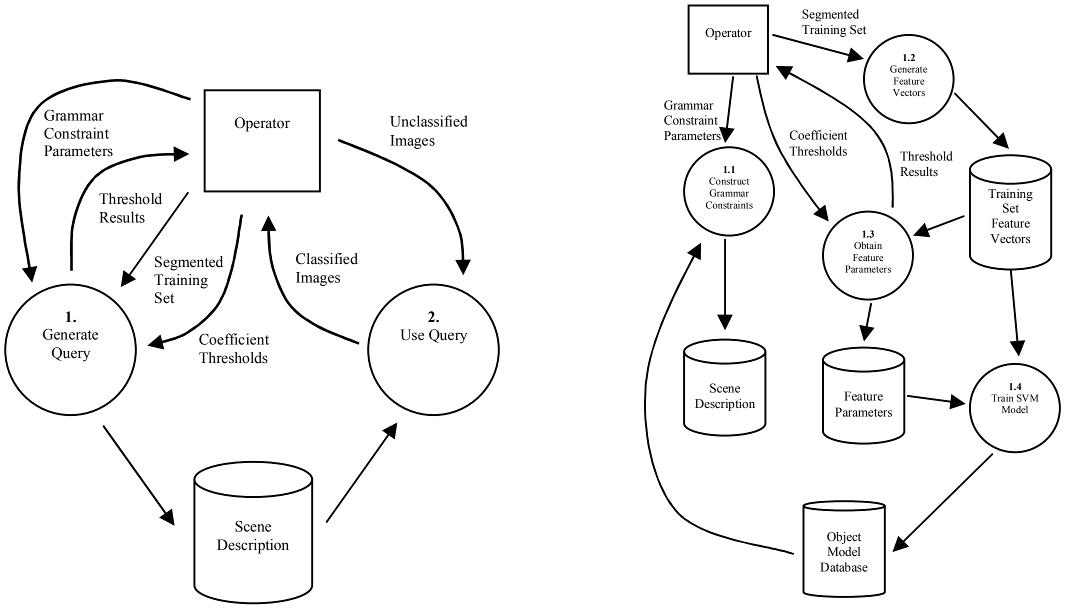

Fig. 1. Overview of the training and querying processes (left) and detail of query generation process (right)

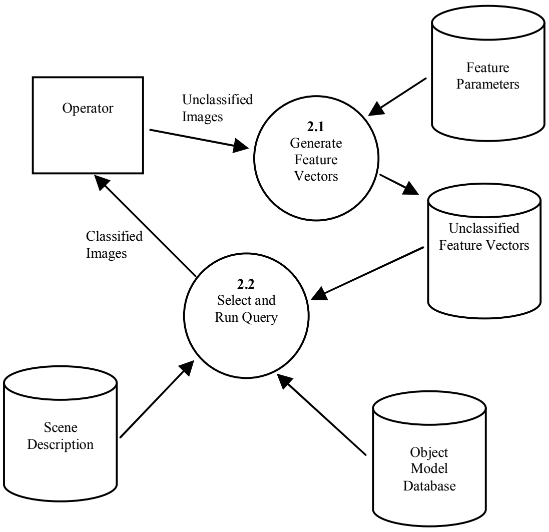

Fig. 3. Detailed diagram of the query usage process

We extend the approach by [17] which used Haar wavelet coefficients as features and SVMs (Support Vector Machine) for training. In their system, the magnitude of the coefficients of two scales (16x16 and 8x8 pixels) and three orientations (horizontal, vertical, diagonal) that indicate the intensity variation are used to locate the position of the components of objects. This multi-level approach is robust and flexible for object configuration design. One drawback is that difficulties due to image scaling and transformations have not been addressed. Our image mining system for computer forensic purposes allows the use of other features (e.g. texture features) in addition to Haar coefficients. We also investigate the effects of using different colour spaces, and of image scaling and transformations. In addition, we examine the needs of effective communication and usage of the system by forensic investigators and relevance feedback for continuous improvement. To this end, we develop a grammar to facilitate the specification of objects, scenes and their relationships. This grammar can also help to filter out invalid configurations. Relevance feedback will be provided via a Bayesian inference network [18].

The image mining module consists of two main parts: training and querying. We separate the two processes because of the differences in technical proficiency and forensic expertise required by each operation. The model trainer sets up parameters used by the classifier and constraints placed on the components of the model in order to train the SVM to recognize certain patches of an image. The query operator runs a query for the classification of a given image, using previously set up queries. Fig. 1 shows the relationships between these two processes.

The training process firstly segments the images in the training set, then calculates feature parameters and obtains appropriate constraints on the model components. These are stored in a database of scene descriptions. A bootstrapping process is then performed until the results are acceptable. This process involves the tweaking of parameters relating to features and constraints, and the retraining of patch detectors after false positive and false negative images from the test runs are added. The output models are stored in an object model database to be used later as query models by the query operator (Fig. 2). In the querying process, the operator supplies an unclassified image. The system segments the image to obtain feature vectors of image patches, then compares them with the models in the object model database and the scene descriptions to obtain a classified image (Fig. 3).

4.1 Performance Results

One application that can benefit from our image mining system is to detect and filter out improper images such as those of partially clad people. We use this application as a case study to test the performance of this system. We use a training set of 214 images consisting of 104 positive images of partially-clad people, and 110 images of negative images of landscapes, textures, clothed people, sport scenes, etc. The patch detectors firstly detect face, waist and pelvis; then combine these components into a hierarchy to detect partially-clad people. Fig. 4 shows a positive image with detected image patches. Each feature vector is composed of high edge coefficients defining the outline of body parts and regions of continuous tones (e.g. bare skin, texture, colour). We perform three experiments using different colour spaces and varying the use of texture homogeneity values. The first test uses HSV space, maximum value of wavelet coefficients in Hue and Value as edge coefficients, and the variance of Hue and Saturation for homogeneous regions. 92% true positive and 74% true negative detection rates are obtained. The second test uses YCbCr space, maximum values of Cb and Cr, and the variance of Cb and Cr. 79% true positive and 95% true negative detection rates are obtained. The third test is similar to the second test except that texture homogeneity values are included as features instead of the variances. The detection rates are the same as in the second test.

4.2 Discussion

From these results, we have found that HSV is more useful for finding positive images, while YCbCr is more discriminating but at a reduced rate of positive detection. The texture homogeneity is not a discriminating feature for this application. Interestingly, we observed that the skin detection using YCbCr has a similar positive rate to that of the SVM classifier. Does this imply that the rate of improvement rests with the choice of a better colour model for skin detection?

5 Grammar-based Specification, Querying and Feedback

To facilitate the communication between forensic investigators and the system, we develop a grammar for describing objects and scenes as hierarchies of component detectors. This grammar defines the position, orientation, error bound, and spatial relationship of the components. Thus, an entire scene can be described as hierarchies at varying levels of resolution, to allow fast search of regions of interest and more detailed and computationally expensive search at a finer level. Users can use this grammar for three tasks: to specify objects and scenes for training, for querying and for providing feedback to the system. Information on the position and orientation is expressed in numerical quantities, while relative spatial arrangement can be expressed in either absolute measurements, or precise terms (e.g. north, south, east, west), or fuzzy terms (e.g. up, down, above, below). These hierarchies which can be represented in an n-ary tree data structure are encapsulated into a file grammar to support storage and manipulation for future use (see Fig. 4).

```
Forensic-Scene:
  Scene
    Scene-Detector-ID
    Comp-Detector-ID
  End-Scene

Scene

  Scene-Detector-ID

  Object-Detector-ID

End-Scene

Comp-Detector:
  Component

    Comp-Detector-ID

    Comp-Detector-Loc

    Displacement$_{opt}$

    Orientation$_{opt}$

    Relation-List$_{opt}$

  End-Component
```

```
Object-Detector:

  Object

    Object-Detector-ID

    Object-Detector-Loc

    Displacement$_{opt}$

    Orientation$_{opt}$

    Relation-List$_{opt}$

    Detector-List

  End-Object

Detector-List:

  Detector-List, Gen-Detector-ID

Gen-Detector-ID: one of
  Object-Detector-ID,

  Comp-Detector-ID,

  Scene-Detector-ID
```

Fig 4. Portion of grammar developed to specify structured image queries.

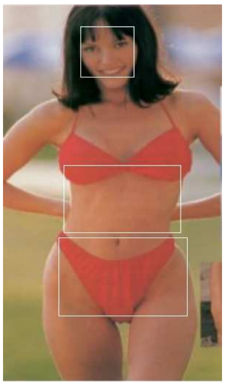

Fig. 4. An example of a positive image

This grammar is extensible to include non-spatial relationships and dynamic scenes. Non-spatial relationships would allow users to specify special characteristics of image evidence based on their previous experience. For example, the co-occurrence of bare skin and pixellated image regions might heighten the chance that the image is pornographic; the co-occurrence of weapons and important buildings might indicate a breach of security. Dynamic scenes occur in motion videos when objects may appear or disappear, or the attributes and relationships between objects may change. These changes can be implemented by appropriate operations on the n-ary tree (insertion, deletion, modification of attributes in the node contents by traversing the tree). Standard transformations (scale, translate, rotate, shear) and linguistic modifications of spatial relationships may be treated as changes in object attributes. To track an object that may be occluded from time to time, a visibility flag is used.

6 Conclusion and Future Work

We have presented a forensic image mining system which is modeled closely to the way forensic investigators work. It provides the facility for training the system to detect the image evidence required, as well as for correcting inaccurate search results or fine-tuning the search further. The communication between users and the system is facilitated by an adaptive grammar. To date, the prototype system consisting of the component-based detection engine and the grammar has been implemented and evaluated for detection of images containing partially clad humans and in other applications with very promising results [6, 7].

The system architecture is flexible in the sense that other types of classifiers (e.g. Naïve Bayes, C4.5 or neural networks) can be used instead of the SVM if they are more suited to the classification of specific types of data. Furthermore, different classifiers may be used for different parts of the system. The grammar is generic and extensible to allow more sophisticated query to be generated if required.

Bayesian networks (BN) provide a compact and efficient means to represent joint distributions over a large number of random variables and allows effective inference from observations (e.g. [18]). Hence, they can be used to understand and learn probabilistic and causal relationships through updating beliefs based on evidence provided. The need for dealing with uncertainties that are inherent in DIF has motivated the use of Bayesian networks. These uncertainties occur in image characteristics, object description, co-occurrence of objects and human semantic interpretation of image content and its relevance to forensic purposes. Our ongoing work includes the implementation of the Bayesian networks for relevance feedbacks and more extensive tests with other examples of image forensic work. It is also envisaged that subjective testing will be performed with input from forensic experts.

References

1. CSI/FBI, 2003 Computer Crime and Security Survey. Computer Security Institute, San Fransisco, USA, 2003.
2. Open Systems, www.opensystems.com.au, visited on 15 Oct. 2003.
3. W. Niblack, X. Zhu, J. Hafner et al, “Updates to the QBIC system”, Storage and Retrieval for Image and Video Databases , 1997, vol. 3312, pp. 150–161.
4. G. Mohay, A. Anderson, B. Collie, O. de Vel and R. McKemmish., Computer and Intrusion Forensics, Artech House Publishers, 2003.
5. H. Muller, W. Muller, S. Marchand-Maillet et al, “Strategies for positive and negative relevance feedback in image retrieval”, Proc. International Conference on Pattern Recognition ICPR2000, vol. 1, pp. 1043–1046.
6. R. Brown, B. Pham and O. de Vel, “A grammar for the specification of forensic image mining searches”, Proc. 8th Australian and New Zealand Conference on Intelligent Information Systems, Sydney, Australia, 2003.
7. Brown, R.; Pham, B., “Image Mining and Retrieval Using Hierarchical Support Vector Machines”, Proc 11th International, Multimedia Modelling Conference (MMM 2005), 12-14 Jan. 2005 pp. 446 – 451.
8. Yow, K. and R. Cipolla, “Feature-based human face detection”, Image and Vision Computing, 1997, vol. 15, (9), pp. 713–735.
9. Haritaoglu, D. Harwood and L. Davis, “W-4: Real-Time surveillance of people and their activities”, IEEE Transactions on Pattern Analysis and Machine Intelligence , 2000, vol. 22, (8), pp. 809–830.
10. Camurri , M. Ricchetti and R. Trocca, “EyesWeb-toward gesture and affect recognition in dance/music interactive systems”, IEEE International Conference on Multimedia Computing and Systems, 1999., Florence, Italy.
11. H. Miyamori, and S. Iisaku , “Video annotation for content-based retrieval using human behavior analysis and domain knowledge”, Proc. Fourth IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000.
12. Tomita, A., T. Echigo, et al., “A visual tracking system for sports video annotation in unconstrained environments” International Conference on Image Processing, Vancouver, Canada, 2000.
13. N. Oliver, B. Rosario and A. Pentland., “A Bayesian computer vision system for modeling human interactions”, IEEE Transactions on Pattern Analysis and Machine Intelligence , 2000, vol. 22, (8), pp. 831–843.
14. Y. Ivanov, and F. Aaron, “Recognition of visual activities and interactions by stochastic parsing”, IEEE Transactions on Pattern Analysis and Machine Intelligence , 2000, vol 22, (8), pp. 852–872.
15. T. Wada and T. Matsuyama, “Multi-object behavior recognition by event-driven selective attention method”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, vol. 22, (8), pp. 873–887.
16. L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition.”, Proceedings of the IEEE., 1989, vol. 77, (2), pp. 257–285.
17. Mohan, C. Papageorgiou and T. Poggio (2001)., “Example-based object detection in images by components”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, vol 23, (4), pp. 349–361.
18. J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Mateo, 1988.