
ABSTRACT

Today a simple search for an image on the internet can return lots of related objects, some are duplicate of the same image and others are unrelated. It is not straightforward to find which image is original from a set of semantically similar images. This paper focus on comparison with traditional near duplicate image detection (NDID) method with image phylogeny tree (IPT) method.

KEYWORDS: Image forensic, multimedia Phylogeny, duplicate de- tection.

INTRODUCTION

In recent years, increasing number of gadget for creating, editing, sharing of digital content brings the problem of handling with a thousands of digital objects (images, or video) who contents are similar. Technological boom of hardware, software and advent of social networks leads to sharing of illegal content and copyright infringement. Most of the forensic investigator faces the problem of finding the original source among a set of related once.

The extensive use of smart phone cameras make legal restrictions on the capture and sharing of digital photographs very difficult. It also helps to alter the digital objects in an impressive manner without need any knowledge or expert.

The researcher in image forensic use traditional near duplicate image detection method to find related images but it give different versions of a given document rather than original image, Image phylogeny structure address the above problem[1][2], which help to find original image from a set of similar images.

This report is organized as follows: Section II presents brief description of near duplicate image detection method, Section III presents image phylogeny tree method to find the original source, Section IV compare the various method and finally, conclusion is presented in Section V.

NEAR DUPLICATE DETECTION METHOD

Finding near duplicate image is a task commonly finds in multimedia information retrieval (MIR) - e.g. detecting illegal copied image in the web. The identification of near duplicate has received a special attention during past few years. It is very common small changes to occur during the retransmission of digital objects without inferring the semantic meaning[4]. For example, once an image or video is shared by a user, it can easily go viral and republished or modified by many other users through web. Modification includes transmission noise

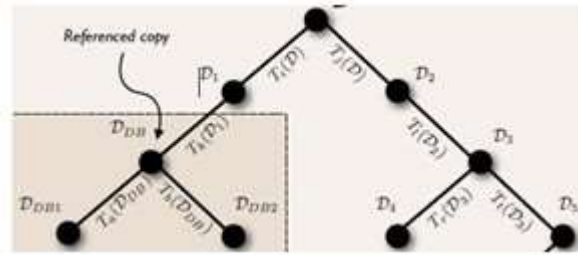


Fig. 1. Example for Near Duplicate Detection

or small editing or corrections such as brightness adjustments or cropping.

When a small changes are applied without disturb the semantic meaning is called near-duplicate object. Near duplicate image detection is defined as follows: Given a set of query images I_q and source image I_s , for each query image find all images that are similar to I_s [6]. There are several applications for NDID technique:[3]

- Reducing the number of versions of a document
- Tracking the legal distribution and spread of a document on the Internet
- Copyright and intellectual property protection
- Multimedia file matching
- Forgery detection

Colour-Based Signature And Matching

It is most widely used method in NDID[6]. Colors are defined on selected color space such as RGB, LAB, LUV, and HSB. A colour histogram simply represents the number of pixels in an image having a particular RGB value. Then, the colour histogram is normalized by dividing by the number of pixels in the histogram. An intersection is applied to the normalized histogram in order to find the similarity between two images.

The term near duplicate or copy refers to any document of the tree. Fig.1 shows that watermarking and fingerprinting methods for the detection of copyright infringement only allows the detection of the copies of the referenced object. On the other hand, content-based approaches enable the detection of all the copies similar images. However, more challenging than identifying different copies of a image is to investigate which image is the original.

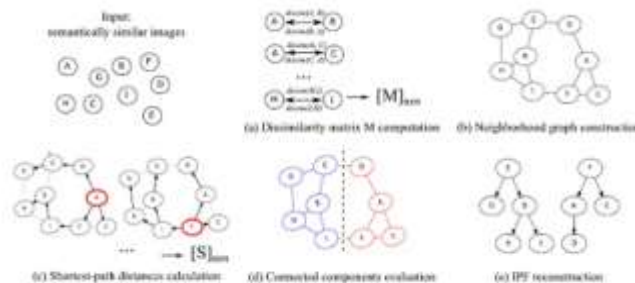


Fig. 2. Steps for Finding Original Image

In this context, a new research field is called multimedia phylogeny is evolved In which we find not only duplicate method but also finding their ancestral relationship and the original source.

IMAGE PHYLOGENY TREE

This method proposes a novel idea by bridging MIR and phylogenic tree in biology, which shows the evolutionary history of organism. Clustering is the important strategy used to separate related and unrelated digital objects[1].

There are several forensic applications for image phylogeny solutions. For example, consider an image on the internet which is redistributed and modified by different users. With phylogenetic algorithm it is easier to understand the ancestral relationship and trace the past history allowing us to find the original image (or the least modified), which can also give hints about the creator of these images. We can solve problem mainly using one tree for set of duplicate images. However, more than one tree is used handling multiple set of near duplicates (ND) in which source image come from different cameras. Approaches for finding phylogeny forests are useful for tracing the original documents within a large set of semantically similar images. In other words, if there is more than one original image, phylogeny forest approaches should be used for finding the correct trees of each image.

IPT for finding original source is describes as follows: if we have n semantically similar images.

- Construct dissimilarity measures between each pair of image.
- Sorting the Dissimilarity value and using a threshold to delimit the number of edges.
- Construct an intermediary graph representing each input as vertices and each edge are dissimilarity values.
- Calculate the shortest path distance between each pair of image. it will generate a new matrix.
- Perform clustering operation afterwards images are correctly grouped.
- Root node represents the original image and edges show the ancestry relationship.

In image phylogeny, the relationships among a set of related digital objects are represented by DAGs, with weights on each directed edge, calculated from the dissimilarity function.

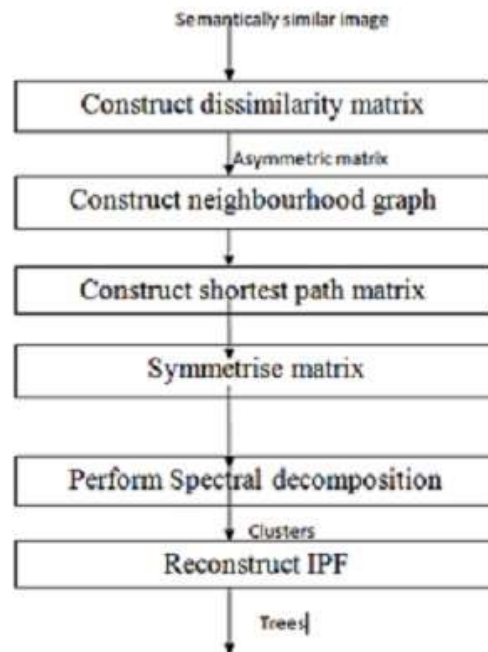


Fig. 3. Flow Chart for Finding Original Source

$$dissim(I_i, I_j) = \min_{T \in \mathcal{L}} |I_j - T_{\beta}^{-1}(I_i)| \text{ point-wise comparison } \mathcal{L},$$

Fig. 4. Capture Function

is the family of transformation . This function measures the amount of residual between the best transformation from image I_i to image I_j , according to T.

Calculating Point Distance Similarities

Considering different data points M that is semantically similar images with pair wise dissimilarities d_{ij} , there are different strategies to construct graph, either using K-NN or neighbourhood graph. In the first approach, an integer k is selected and each point has exactly k neighbors, which are the k closest points to itself; in the second approach, a real number is selected, and G is constructed by connecting the neighbour.

COMPARISON

Compared to the two techniques used in multimedia image retrieval, near duplicate image detection method only find duplicate of an image. However in phylogeny tree problem address

- 1) Duplicate Detection
- 2) Trace Ancestral Relationship
- 3) Find the original source image

So the IPT is the relevant method to solve current image forensic problems.

Table I Comparison OF NDID and IPT

Sno.	Method	Functions
1	NDID	<ul style="list-style-type: none"> • Only give different versions of a image
2	IPT	<ul style="list-style-type: none"> • Duplicate detection • Find original source • Trace history

CONCLUSION AND FUTURE WORK

Image Phylogeny is a research field with several applications in digital forensic, which help investigators to find original image or illegal content from a set of related ones.

Compared to traditional NDID problems IPT address wide range solutions and applications. In future work is focus to propose a novel method for finding who published the original image blending with multimedia phylogeny

REFERENCES

- [1] Z. Dias, A. Rocha, and S. Goldenstein :First steps toward image phylogeny, in Proc. IEEE Int. Workshop Inf. Forensics Security, Dec 2010,
- [2] Z. Dias, S. Goldenstein, and A. Rocha, Large-scale image phylogeny:Tracing image ancestral relationships, IEEE Multimedia, Jul./Sep. 2013.
- [3] Z. Dias, S. Goldenstein, and A. Rocha, Toward image phylogeny forests: Automatically recovering semantically similar image relationships, Forensic Sci. , 2013.
- [4] F. D. O. Costa, M. A. Oikawa, Z. Dias, S. Goldenstein, and A. Rezende de Rocha, Image phylogeny forests reconstruction, IEEE Trans. Inf. Forensics Security Oct 2014.
- [5] Z. Dias, A. Rocha, and S. Goldenstein, First steps toward image phylogeny, in Proc. IEEE Int. Workshop Inf. Forensics Security, Dec. 2010.
- [6] D. Cai, X. He, and J. Han. Spectral Regression: A Unified Subspace Learning Framework for Content-Based Image Retrieval. In ACM Multimedia, 2007.