

Towards Detecting Faked Images

Khaled A. N. Rashed and Ralf Klamma

Databases & Information Systems
RWTH Aachen University, Ahornstr. 55, D-52056 Aachen, Germany
`rashed,klammas@dbis.rwth-aachen.de`

Abstract. An important facet of traditional retrieval models is that they retrieve images and videos and consider their content and context reliable. Nevertheless, this consideration is no longer valid since they can be faked for many reasons and at different degrees thanks to powerful multimedia manipulation software. Our goal is to investigate new ways detecting possible fake in social network platforms. In this paper, we propose an approach that assets identification faked images by combining standard content-based image retrieval (CBIR) techniques and community as alternative solution for supporting such semantic multimedia tasks in an interoperable way using ontologies.

1 Introduction

Nowadays, in the Web 2.0 platforms, there has been an explosive growth of multimedia content. One of the results of the dynamic nature of multimedia content in social network sites is the fake of images. By image fake we mean the changing of an image's context or its content or both of them for whatever reason. Identification of such fake is considered an important issue for a number of applications such as digital cultural heritage, eTourism, eLearning and multimedia management in general. Identification of faked images in large scale social network sites such as Flickr¹ remains an elusive goal. Over the last years, a number of approaches have tackled the underlying task such as digital watermarking approaches [7] that for a long time have been proposed to provide image authenticity. These approaches rely on embedding additional information within the image content. Their limitation is the requirement of embedding information in images before making them public, which excludes images that are already in public platforms. In addition, watermarking modifies the image content that might create a serious problem where the quality is highly required. Digital forensics tools [6] aim to determine whether digital photos have been manipulated. These tools can measure statistical inconsistencies in the underlying image pixels, detecting traces of resampling, improbable lighting and shadow, physically distortion, and other artifacts. Disadvantage of such tools is that their use in public domains is computationally expensive. Content based approaches such as replica detection [5], near-duplicate detection [3] aim at detecting all

¹ <http://flickr.com>

images reproduced from the original image. They are based on the image similarity. Their response speed and efficiency of such detection scheme is largely affected by the size of the original/reference image dataset. The main advantage of such approaches stems from the fact that no additional information should be embedded within the image, as they consider the image itself is the watermark. Content-based image retrieval is the main element of these systems. In addition, it has the scalability property that makes it a suitable alternative to be used to tackle the fake detection problem in social network platforms. The amount of images uploaded to the social network sites to be analyzed by other communities grows exponentially. Therefore, developing new techniques, that can help to make a statement about the searched or retrieved image is strongly needed. Our goal is to combine semantic multimedia metadata, communities and Web 2.0 with content based image retrieval to solve the fake images problem.

CBIR Techniques for Faked Image Detection

The success of utilizing content-based image retrieval (CBIR) techniques [2] in widely application areas motivates us to use them to manage low-level image features. In this regard, images are represented in the feature vector space by means of MPEG-7 visual descriptors [4]. MPEG-7 standard facilitates the quality access to the image content, and support the interoperability issue. Furthermore, MPEG-7 captures the different aspects of color, texture and shapes of the images. They are also robust against the common image transformations. Similarity between two images is measured by calculating the distances between two feature vectors using standard functions.

Web 2.0, Community and Collaborative Activities

Web 2.0 refers to the infrastructures that support humans working together by means of tools and technologies. Community approach in the context of fake detection in such platforms focuses on detecting semantic inconsistencies in images. Rating plays an important role in influencing decision making and people's choices. Therefore, we will use the collaborative rating to images against the suspected fake. Collaborative rating in Web 2.0 allows classification of experts in dynamic manner. One of most important problems in this regard is the trust of individual users and the quality of ratings. Users rate images and create the associations between the users, rates and the images. Collaborative filtering is used to automatically identify high quality rates for users. Collaborative rating network can be modeled as social network. Such network presents the graph with ternary edges, where each edge represents as the fact that a particular user rated certain image with certain rate. Trust can be calculated among network users based on their dynamic relationships.

2 Objectives and Approach

We intend to develop a system that enables to make sense about the images fake. Such system needs to be able to manage the image content as well as semantic users' interactions. We believe that content-based image retrieval (CBIR) tech-

niques and harnessing collective intelligence² is the appropriate solution. In our work, we consider the image as faked if some details are deleted from or inserted to it and photomontage. Another types of image transformation (i.e. rotation, smoothing, luminance change, compression and additive noise) are accounted as versions of the original image. Since a faked image is generated based on an original image, the requirements of the image fakery is that it retains similar visual content to the original image. Therefore, the faked image shares lots of information with the original enough to be distinguished from any other images. Based on this fact, detecting faked images requires robust and efficient visual features to be managed. To fulfill the above requirements and to achieve the interoperability, well established MPEG-7 standards and tools [8] will be used, MPEG-7 is able to express image content covering the most aspects including low level technical information and high level content semantics. Among many useful MPEG-7 visual descriptors, we will consider: *Color Layout*, *Scalable Color*, *Homogeneous Texture*, *Edge Histogram* and *Color Structure* to create the image signature. The low level image features do not provide a comprehensive infor-

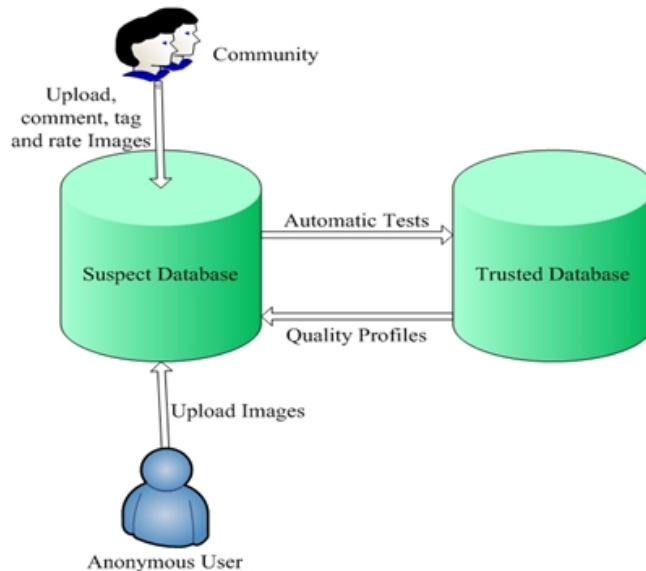


Fig. 1. Overview of the System

mation about the image fake, but they can be support the community decision. On the other hand, we also intent to deploy community and Web 2.0 aspects in our fake detection approach. The key idea is behind the fact that most of discovered faked images were identified by some one who knows the original and then

² <http://www.oreilly.de/artikel/web20.html>

those images have been analyzed against the fake. The overview of the proposed approach is visually described in the fig.(1). Normal users and community can upload images to the system from social network sites such Flickr. These images can be peer rated and commented in term of the fake (degree of fake). For each uploaded image extracted the image signature, and will be stored together with corresponded rates and possible tags in suspected database that should be publicly accessed. Trusted database stores all information about original images, including MPEG-7 visual descriptors, tags, ratings and comments. In case of existing reference trusted image, low level features will be compared, the results will be visualized and then the user judgment will be taken by rate and tag of suspected image. This procedure will be happened reciprocally. Evidence from the community knowledge can be complemented by the evidence from the low level features and vice versa. Information from the trusted database used to improve quality of rating in the case of existing the reference image. When no reference image existed we rely only on the community knowledge. Users (community members) look at image and try to find out the semantic errors that can be exist in the image, annotate and rate the image.

3 Conclusions

In this paper, we presented an overview of our proposed approach to detect faked images. We believe that with combining CBIR with emerging Web 2.0 concepts such as collective intelligence and collaborative rating the fake detection problem in the large scale community image sharing platforms can be solved. Prototype to prove the concepts described above will be implemented. Ontologies can serve as the basis to combine two mention above approaches. In the future work we plan also to investigate on more types of fake issues. Considering image signature tools [1]. We also intend to test some combinations of different feature sets, not only to detect the manipulated images but also to detect the similar fake.

References

1. "ISO/IEC 15938-3:2002/Amd.3 Image Signature Tools ", Mar 2009.
2. R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of new age. *ACM Computing Surveys (CSUR)*, 40:60, 2008.
3. Y. Ke, R. Sukthankar, L. Huston, Y. Ke, and R. Sukthankar. Efficient near-duplicate detection and sub-image retrieval. In *In ACM Multimedia*, pages 869–876, 2004.
4. MPEG-7. Multimedia content description interface part 3: Visual. *ISO/IEC JTC1/SC29/WG11,Doc.N4062*, Mar 2001.
5. S. Nikolopoulos, S. Zafeiriou, and N. Nikolaidis. Image replica detection system utilizing r-trees and linear discriminant analysis. *PR*, 43(3):636–649, March 2010.
6. A. C. Popescu and H. Farid. Exposing digital forgeries by detecting traces of re-sampling. *IEEE Transaction on signal processing* 53, 53(758-767), 2005.
7. C. Rey and J.-L. Dugelay. A survey of watermarking algorithms for image authentication. *EURASIP J. Appl. Signal Process.*, 2002(1):613–621, 2002.
8. G. Shih-Fu Chang, A. B. Chan, and P. J. Moreno. Overview of the mpeg-7 standard. *IEEE Trans. Circuits and Systems for Video Technology.*, 11(0):688 – 695, 2001.