Towards efficient privacy-preserving image similarity detection

Cite as: AIP Conference Proceedings **2144**, 050005 (2019); https://doi.org/10.1063/1.5123121
Published Online: 23 August 2019

Ayad I. Abdulsada, and Nagham Abdulrasool Taha

Towards efficient Privacy- Preserving Image Similarity Detection

Ayad I. Abdulsada[1, a)] and Nagham Abdulrasool Taha [1,b)]

[1]*Universiyt of Basrah, Education College for pure Science Basrah, Iraq.*

[a)]Corresponding author: ayad.abdulsada@uobasrah.edu.iq
[b)]naghamrsool@gmail.com

Abstract. Image similarity plays an essential role in many real world applications such as content based image retrieval (CBIR), computer vision, and near duplicate image detection. The state of the art methods are generally assumed that the content of images is not private. This reduces the utilization of these methods to work within only environments where images are publicly access. Essentially, this assumption limits more practical applications, e.g., image matching between two security agencies, where images are confidential. We address the problem of evaluating the similarity between image collections of two parties, who are reluctant to reveal their actual content. The Euclidian distance measure is used to measure distances between global hierarchal color histograms. We conduct several experiments on real image collections to demonstrate the practical value of the proposed schemes.

INTRODUCTION

Image similarity is ubiquitous in many real-word applications. For example, near duplicate image detection [1], content based image retrieval [2], biometric matching, and file management. Advances in object recognition make the use of biometric data more prevalent method of authentication and identification. Particularly, several security agencies around the world collect biometric features from travelers. The rapid advances in digital cameras and media-enabled smartphones, led to generate numerous high-resolution photos every day, which makes storing, managing and searching large-scale images become challenging task. For this reason, it is desirable to offer a secure management service (and storage) for large scale multimedia data. Tagging all uploaded data manually for enabling a keyword based search is unpractical task. To simplify such a process, researchers focused on developing tools to analyze and extract features from the content of digital media automatically. One example of such tools is content based image retrieval (CBIR), where features are extracted from images, then features of different images are compared to show how similar they are.

Most similar image detection approaches work by defining distance measure and image representation that reduce the amount of data stored per image and the time cost of database search. The amount of stored data could be either a single global color histogram per image or large sets of feature descriptors e.g. SIFT[3] and SURF[4]. The first method generates an extreme compressed feature vector for each image. Such method is effective to identify global similarities, e.g., how many colors two images share. The second method searches the image to identify interest key points that are invariant to scale and orientation. Then for each key point one feature descriptor is extracted to describe the visual characteristics around this point. For a specific type of descriptor, feature vectors are usually of the same dimension, e.g., the feature vector of SIFT descriptor is 128-dimension. When searching the database for similar image, algorithms of different time complexity are proposed, the simplest method being computing the similarity to every image in the database. Similarity in images is often expressed as the similarity in their descriptors. If two documents have similar descriptors, they are marked as similar. In this paper, we employ this model to capture similarity between images.

While image similarity increasingly employed as a good mechanism for detecting similarity of objects, it is still undesirable for applications of sensitive data that must be subject to minimal exposure. This leads to develop new

The 7th International Conference on Applied Science and Technology (ICAST 2019)
AIP Conf. Proc. 2144, 050005-1–050005-12; https://doi.org/10.1063/1.5123121
Published by AIP Publishing. 978-0-7354-1889-9/$30.00