IGI PUBLISHING
ITJ 5041
701 E. Chocolate Avenue, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.igi-global.com

This paper appears in the publication, International Journal of Digital Crime and Forensics, Volume 1, Issue 3
edited by **Chang-Tsun Li © 2009, IGI Global**

Robust Near Duplicate Image Matching for Digital Image Forensics

H.R. Chennamma, University of Mysore, India
Lalitha Rangarajan, University of Mysore, India
M.S. Rao, Indian Academy of Forensic Sciences, India

ABSTRACT

Local invariant key point extraction has recently emerged as an attractive approach for detecting near duplicate images. Near duplicate images can be: (i) perceptually identical images (e.g. allowing for change in color balance, change in brightness, compression artifacts, contrast adjustment, rotation, cropping, filtering, scaling etc.), (ii) images of the same 3D scene (from different viewpoints). The requirements for identifying near duplicate images vary according to the application. In this paper we focus on image matching strategy that will assist in the detection of forged (copy-paste forgery) images. So far, no specific image matching strategy exists for this application. The state of the art methodologies tend to generate many false positives. In this paper we have introduced a novel matching strategy for pattern matching of key point distributions. Typical experiments conducted with real world images demonstrate success in near duplicate image retrieval for the application of digital image forensics. Proposed method outperforms some of the existing methods and is computationally efficient. [Article copies are available for purchase from InfoSci-on-Demand.com]

Keywords: *Angle and Line Ratio Statistics; Detecting Copyright Violations; Local Invariant Features; Near Duplicate Image Matching; Spatial Consistency*

INTRODUCTION

Forensic experts believe that no criminal can do his activities without leaving evidence at the scene of crime. However it is very difficult to trace out evidences especially in case of digital image forgeries. Content based near duplicate image matching has recently emerged as a new approach for identifying suspicious pirated copies of digital images. Local invariant feature based approaches for image matching has been successfully applied to a wide range of problems including object recognition (Lowe, 2004), panoramic image stitching (Brown & Lowe, 2003), image mosaicing (Qi & Jeremy, 2006), near duplicate image detection (Zhao, 2007) etc. Near duplicate

image detection and retrieval is a vital component for many real world applications and recently is being used for news story threading (Chang, 2005), content based video search (Chang, 2005), topic detection & tracking (Wu, 2006), near duplicate shot detection in video (Ondrej, 2007) and copyright infringement detection (Ke & Sukthankar, 2004b). Effective and reliable watermarking algorithms for detecting copyright infringement are not yet available. Content based near duplicate image matching can be a complementary approach in the identification of image ownership.

The definition of a Near Duplicate Image (NDI) varies depending on what photometric and geometric variations are deemed acceptable and the application in hand. In the case of exact duplicate detection, no changes are allowed. At the other extreme, a more general definition is that these are images of the same scene but with possibly different viewpoints or the perceptually identical images which are slightly altered versions (using powerful image processing tool) of its original. In this article we mainly focus on image matching strategy for finding copies, fragments of images or variants of the given suspicious digital image(copy-paste forgery). So we refer images as NDIs if these images are perceptually identical but not recognized as such due to common image manipulations such as change in color balance, change in brightness, change in file formats, compression artifacts, contrast adjustment, rotation, cropping, filtering, scaling etc. As we are searching for the images which are altered versions of the original image, the images with slight viewpoint variations of the same scene (called as similar images) are not likely to be expected.

Many matching techniques that use invariant local features extract key points from all images and then the query image features vote independently for features from the database images (votes are computed based on proximity and similarity of their intensity neighborhood). The greater the number of votes found, the more likely it is that the image is near duplicate. However, it is still likely that there are significant false positives at the key point matching phase. In other words, although some key points are within the threshold distance, they belong to patches of images that are not near duplicates. So it must be followed by a verification step to account for spatial or geometric relationships between the extracted key points. We need an image matching strategy which is robust enough to decide inliers and outliers according to the application.

Unfortunately the existing matching techniques fail to distinguish similar images of the same scene from the original image which has been used in the creation of fake image. Hence the current state of the art image matching methods end with many false positives (images that are not near duplicates of the query image). In order to overcome this drawback, we propose a novel pattern matching technique for finding copies, fragments of images or variants of the same digital image. The proposed NDI identification system is useful when the copyrighted images are stored in a system. We could then detect query images that were composites and accurately identify the exact sources used in their creation.

Rest of this article is organized as follows. Section 2 reviews the relevant research. Section 3 gives an overview of the near duplicate image detection system in two steps. Section 4 describes our proposed algorithm for indexing feature points and