Neapolis University

HEPHAESTUS Repository **http://hephaestus.nup.ac.cy**

School of Information Sciences Conference papers

2014

Searching Images with MPEG-7 (& MPEG-7-like) Powered Localized dEscriptors: The SIMPLE answer to effective Content Based Image Retrieval

Iakovidou, Chrysanthi

IEEE (Computer Society)

http://hdl.handle.net/11728/10213

Downloaded from HEPHAESTUS Repository, Neapolis University institutional repository

Searching Images with MPEG-7 (& MPEG-7-like) Powered Localized dEscriptors: The SIMPLE answer to effective Content Based Image Retrieval

Chryssanthi Iakovidou, Nektarios Anagnostopoulos, Athanasios Ch. Kapoutsis,
Yiannis Boutalis and Savvas A. Chatzichristofis
Department of Electrical and Computer Engineering
Democritus University of Thrace
(DUTH), Xanthi, Greece
{ciakovid,nektanag,akapouts,ybout,schatzic}@ee.duth.gr

***Abstract*—In this paper we propose and evaluate a new technique that localizes the description ability of the well established MPEG-7 and MPEG-7-like global descriptors. We employ the SURF detector to define salient image patches of blob-like textures and use the MPEG-7 Scalable Color (SC), Color Layout (CL) and Edge Histogram (EH) descriptors and the global MPEG-7-like Color and Edge Directivity Descriptor (CEDD), to produce the final local features' vectors. In order to test the new descriptors in the most straightforward fashion, we use the Bag-Of-Visual-Words framework for indexing and retrieval. The experimental results conducted on two different benchmark databases with varying codebook sizes, revealed an astonishing boost in the retrieval performance of the proposed descriptors compared both to their own performance (in their original form) and to other state-of-the-art methods of local and global descriptors. Open-source implementation of the proposed descriptors is available in c#, Java and MATLAB[1].**

I. Introduction

After many years of research, little is known about the combination of features that best describes an image with respect to its visual properties or its visual content. With image collections growing by the minute in various areas such as medicine, private life, industrial/commercial products, journalism, tourist attractions and art -to name a few- a plethora of Content Based Image Retrieval (CBIR) systems have been introduced in the literature. The main objective of all proposed schemes is to represent images with a feature vector or descriptor that will allow fast access and meaningful retrieval for the user.

Representing images with numerical values in a way that grasps their distinctive visual properties and contents, is a challenging process. The variety of solutions and proposed implementations of description methods in the literature is indicative of the complexity of the problem. A wide collection of early strategies and recent trends on image retrieval can be found in the well-structured studies [1], [2], while feature specific studies evaluating color description [3]–[5], texture description [6]–[8] and shape description [9], [10] strategies, vividly outline the many directions researchers explored in the quest of effectively representing image content.

In essence, the success of any CBIR system is subject to the user's requirements and the specific characteristics of the image collection. When a query is set with the objective to retrieve visually similar images, for instance natural scenes of mountains or fields and forests, or even images depicting objects with little or no background clutter, a feature vector that treats and describes the images as a whole, is effective [11]. Global Features (GF) such as color, texture and shape are calculated on the entire image to form an informative feature vector representation. The images in a collection are compared with various distance measures to the query. The lower the distance, the higher the rank they achieve in the retrieval process.

On the other side of the spectrum, when searching for images with similar visual and conceptual content, which is the case for verbose images or images where objects appear with partial occlusions [12], global vectors fail to discriminate the constituent parts of an image. For example, objects of the same shape described by a global image vector will not be distinguished even if their local texture information varies significantly. The information concerning localized aspects of an image can be of great importance for representing and classifying images that present high in-class variability.

In order to enrich the representation with a local information component, Local Features (LF) were introduced. In theory, every pixel in an image can be used to define a LF. This would lead to an unmanageable number of features that do not necessarily add to the descriptor's discrimination ability. Thus, LF are vector representations of salient regions of the image. These salient regions, often referred to as points-of-interest (POI), are local extrema of some function of the image, like edges, corners and blobs. Some among the most widely employed POI detectors are corner detectors Harris [13], Shi-Tomasi [14], and FAST [15] and blob detectors SIFT [16], SURF [17], to name a few. In [18] the authors provide some basic guidelines concerning the proper selection of detectors, as they investigate their pros and cons for usage in visual odometry. An overview of the detectors can be found in [19].

[1] http://tinyurl.com/SIMPLE-Descriptors

978-1-4799-3990-9/14/$31.00 ©2014 IEEE