This full text paper was peer-reviewed at the direction of IEEE Instrumentation and Measurement Society prior to the acceptance and publication.

Localized Global Descriptors for Image Retrieval: An Extensive Evaluation on Adaptations to the Simple Model

N. Anagnostopoulos[1], C. Iakovidou[2], S. A. Chatzichristofis[2], Y. Boutalis[2] and M. Lux[1]
[1]Institute of Information Technology, Klagenfurt University,[2] Democritus University of Thrace, Greece
nektarios.anagnostopoulos@aau.at, ciakovid@ee.duth.gr, schatzic@ee.duth.gr, ybout@ee.duth.gr, mlux@itec.aau.at

***Index Terms*—Content Based Image Retrieval, Local Features, Bag of Visual Words, VLAD**

***Abstract*—SIMPLE (Searching Images with MPEG-7 (& MPEG-7-like) Powered Localized dEscriptors) is a model that proposes the reuse of well-established global descriptors by localizing their description mechanism on image patches located by local features' detectors. Having displayed impressive retrieval results on two different databases, in this paper we extend the family by replacing the originally picked global descriptors and by applying VLAD, a BOVW alternative, for the vectorization of the features. We re-evaluate all SIMPLE descriptors (original and proposed here) on five benchmarking databases featuring very diverse scenarios, so as to investigate how SIMPLE works with different image retrieval cases, ranging from near duplicate search to visual object similarity. The experimental results show the robustness of the scheme with SIMPLE descriptors presenting stable and high retrieval performances across all tested collections, outperforming not only the methods they originated from, but also some of the best reported state-of-the-art methods as well.**

I. Introduction

SIMPLE descriptors, first introduced in [1], [2], are produced by employing a key-point detector (SURF [3]) to define salient image patches, which are then described using a global descriptor from the literature. The global descriptors originally used were the MPEG-7 SCD, CLD and EHD [4] and the MPEG-7-like CEDD [5] descriptors. The global descriptors employed to be localized are all compact and quantized in order to be practical for CBIR. For the aggregation of the features the Bag-Of-Visual-Words (BOVW) [6] architecture was implemented. The minimum requirement for a global descriptor to become a SIMPLE descriptor is to be compact and quantized.

In this paper we extend the SIMPLE family by varying the description method on the extracted patches and the aggregation method of the produced features. We employ two descriptors from the literature that are compact/composite/quantized and include color information (FCTH[7] and JCD[8]) and are thus expected to be perfect candidates for localization through SIMPLE. We also go on and employ two additional descriptors that partially meet the *perfect candidate* criteria: (i) the Opponent Histogram [9], a compact and quantized color descriptor, and (ii) the Auto Color Correlograms [10], a composite, quantized (but not compact) color descriptor. For the features' aggregation we test both BOVW and a recently proposed alternative, the Vector of Locally Aggregated Descriptors (VLAD) [11]. We evaluate the scheme on five different benchmarking image collections with the goal to investigate the robustness of SIMPLE over different types of retrieval scenarios and to minimize the case that good achieved performances might have to do with specificities of the database. The retrieval performances of the new SIMPLE descriptors are compared to the global descriptors they originated from, as well as methods reported in literature.

II. The SIMPLE Model

Four stages are involved in building a SIMPLE descriptor: First, the SURF detector is employed to locate Points-of-Interest (POIs) in images. These POIs are salient spots in the image considered containing blob-like structures in multiple scales, whose size is determined as a square area that depends on the scale that the points were detected in.

Next, a global descriptor from the literature that is compact and quantized is applied on those patches as if they were standalone images. Originally, three MPEG-7 descriptors (SCD, CLD, EHD) and an MPEG-7-like descriptor (CEDD) were used to describe the patches and produce a respective number of different SIMPLE local features (SIMPLE-SC, SIMPLE-CL, SIMPLE-EH and SIMPLE-CEDD).

A random sample of SIMPLE features, extracted from the whole image collection, is forwarded to a k-means classifier to produce codebooks of predefined numbers of clusters, aka *visual words*. The BOVW model is used to construct the vector representation of the images. Moreover, eight different weighting schemes can be applied. Results are obtained by the Euclidean L_2 distance between the query's vector and the vectors of the images in the collection.

SIMPLE descriptors were extensively evaluated in [1] by performing multiple experiments on two well-known benchmarking databases (UCID [12] and UkBench [13]). The experimental results, showed that the hybrid approach that the SIMPLE method implements, boosts the retrieval performance of the descriptors both compared to their original global form and to other state-of-the-art descriptors from the literature (such as SIFT [14], SURF[3], BRISK[15], ORB[16] and Opponent SIFT[17]). Particularly impressive results are achieved when the employed description method carries color

978-1-5090-1817-8/16/$31.00 ©2016 IEEE