Accepted Manuscript

Object-oriented convolutional features for fine-grained image retrieval in large surveillance datasets

Jamil Ahmad, Khan Muhammad, Sambit Bakshi, Sung Wook Baik

PII: S0167-739X(17)31857-5
DOI: https://doi.org/10.1016/j.future.2017.11.002
Reference: FUTURE 3794

To appear in: *Future Generation Computer Systems*

Received date : 16 August 2017
Revised date : 18 October 2017
Accepted date : 1 November 2017

Please cite this article as: J. Ahmad, K. Muhammad, S. Bakshi, S.W. Baik, Object-oriented convolutional features for fine-grained image retrieval in large surveillance datasets, *Future Generation Computer Systems* (2017), https://doi.org/10.1016/j.future.2017.11.002

Object-Oriented Convolutional Features for Fine-Grained Image Retrieval in Large Surveillance Datasets

[1]Jamil Ahmad, [1]Khan Muhammad, [2]Sambit Bakshi, [1,*]Sung Wook Baik
[1] Digital Contents Research Institute, Sejong University, Seoul, Republic of Korea
[2] Department of Computer Science & Engineering, National Institute of Technology, Rourkela, India
[*]Corresponding author: sbaik3797p@gmail.com

Abstract

Large scale visual surveillance generates huge volumes of data at a rapid pace, giving rise to massive image repositories. Efficient and reliable access to relevant data in these ever growing databases is a highly challenging task due to the complex nature of surveillance objects. Furthermore, inter-class visual similarity between vehicles requires extraction of fine-grained and highly discriminative features. In recent years, features from deep convolutional neural networks (CNN) have exhibited state-of-the-art performance in image retrieval. However, these features have been used without regard to their sensitivity to objects of a particular class. In this paper, we propose an object-oriented feature selection mechanism for deep convolutional features from a pre-trained CNN. Convolutional feature maps from a deep layer are selected based on the analysis of their responses to surveillance objects. The selected features serve to represent semantic features of surveillance objects and their parts with minimal influence of the background, effectively eliminating the need for background removal procedure prior to features extraction. Layer-wise mean activations from the selected features maps form the discriminative descriptor for each object. These object-oriented convolutional features (OOCF) are then projected onto low-dimensional hamming space using locality sensitive hashing approaches. The resulting compact binary hash codes allow efficient retrieval within large scale datasets. Results on five challenging datasets reveal that OOCF achieves better precision and recall than the full feature set for objects with varying backgrounds.

Keywords: image retrieval, object-oriented features, convolutional neural network, fine-grained retrieval

1. Introduction

In recent years, we have seen tremendous increase in the production and consumption of multimedia data partly due to advent of the social web and partly because of the progress in surveillance, medical, industrial, mobile and embedded computing technologies [1]. Consequently, multimedia data including images and videos are produced and stored in huge amounts. These multimedia repositories contain wealth of highly useful information for administrators and decision makers, provided that efficient and reliable access to relevant data is ensured [2]. Content-based image retrieval (CBIR) systems attempt to locate images containing objects similar to that of a query image by analyzing their contents. CBIR has several applications in information retrieval, surveillance, medical, e-commerce, industry, and social web. Recently, it has attracted a lot of attention due to the rising interest in making the best use of available multimedia data [3]. The exponential increase in the volume of image data, and the inherent complexity of visual contents (projecting 3D world onto a 2D canvas) has made image retrieval increasingly difficult. This difficulty increases even further with fine-grained image retrieval due to the existence of high degree inter-class visual similarity [4]. One such problem arises when retrieving images from traffic surveillance datasets, where the main object of interest are vehicles [5, 6]. There exists greater visual similarity despite the fact that vehicles may belong to different categories.