Electronic Letters on Computer Vision and Image Analysis 16(2):5-8, 2017

Selection of relevant information to improve Image Classification using Bag of Visual Words

Eduardo Fidalgo Fernández

Dep. of Elect., Systems and Automatic Engineering, Universidad de León, Campus de Vegazana S/N, 24071, León, Spain
Researcher at INCIBE (Spanish National Institute of Cybersecurity), León, Spain

Advisor/s: Enrique Alegre and Victor González-Castro
Date and location of PhD thesis defense: 9 December 2015, Universidad de León
Received 29th Jun 2017; Accepted 20th Sep 2017

Abstract

In the problems where there is not enough data to use Deep Learning, traditional techniques like Bag of Visual Words (BoVW) together with the use of hand-crafted features still obtain outstanding results. In this work, we introduced several contributions in the feature extraction stage of the BoVW framework. We first evaluated how the early fusion of SIFT and Edge-SIFT descriptors affects the classification performance, demonstrating that the parameter to get the Edge-SIFT descriptors proposed in the literature, i.e. radius, is not the optimum one for an image classification task. Following this research line, we proposed Compass Radius Estimation for Image Classification (CREIC), a new method to estimate a radius to compute Edge-SIFT descriptors on a dataset. On a separate research line, We introduced the Saliency maps into the BoVW framework, due to their efficiency in pointing out the attention on relevant parts of an image. At this stage, we proposed three Semantic Attention Filtering strategies; one at Keypoint level based on dictionary distances (SAKF) and two at region level based on Intersection of Saliency Maps (SARF-ISM) and Keypoint Voting (SARF-KV). These approaches remove features which mainly belongs to the image background, resulting in an accuracy increase in an Image Classification process. Finally, considering a Saliency Map as a topographic map, we demonstrated that the information extracted at different levels of this map affects the image classification results. With a Semantic Weighted Schema (SWS), we obtained various levels of information from an image, and we evaluated their contribution and combination concerning the accuracy obtained with different trained models.

1 Introduction and Motivation

One of the main challenges in computer vision is image classification. Nowadays the number of images increases exponentially every day; therefore, it is important to classify them reliably. The conventional image classification pipeline usually consists on extracting local image features, encoding them as a feature vector and classify them using a previously created model. Without taking into account Deep Learning models, in feature codification, the Bag of visual Words (BoVW) model and its extensions, such as pyramid matching and weighted schemes, achieved quite good results in the last years and have became the state of the art methods when the number of images is not high.

Correspondence to: <efidf@unileon.es>

Recommended for acceptance by Anjan Dutta and Carles Sánchez
https://doi.org/10.5565/rev/elcvia.1102
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

The process as mentioned above is not perfect and computers, as well as humans, may make mistakes in any of the steps, causing a drop in classification performance. Some of the primary sources of error on large-scale image classification are the presence of multiple objects in the image, small or skinny objects, incorrect annotations or fine-grained recognition tasks among others.

Based on those problems and the steps of a typical image classification pipeline, the motivation of this work was to provide some guidelines to improve the quality of the extracted features to obtain better classification results. The contributions of this work demonstrate how a good feature selection can contribute to improving the fine-grained classification, and that there would even be no need to have a big training data set to learn the key features of each class and to predict with good results. Figure 1 introduces the main contributions of this work in the feature extraction phase.

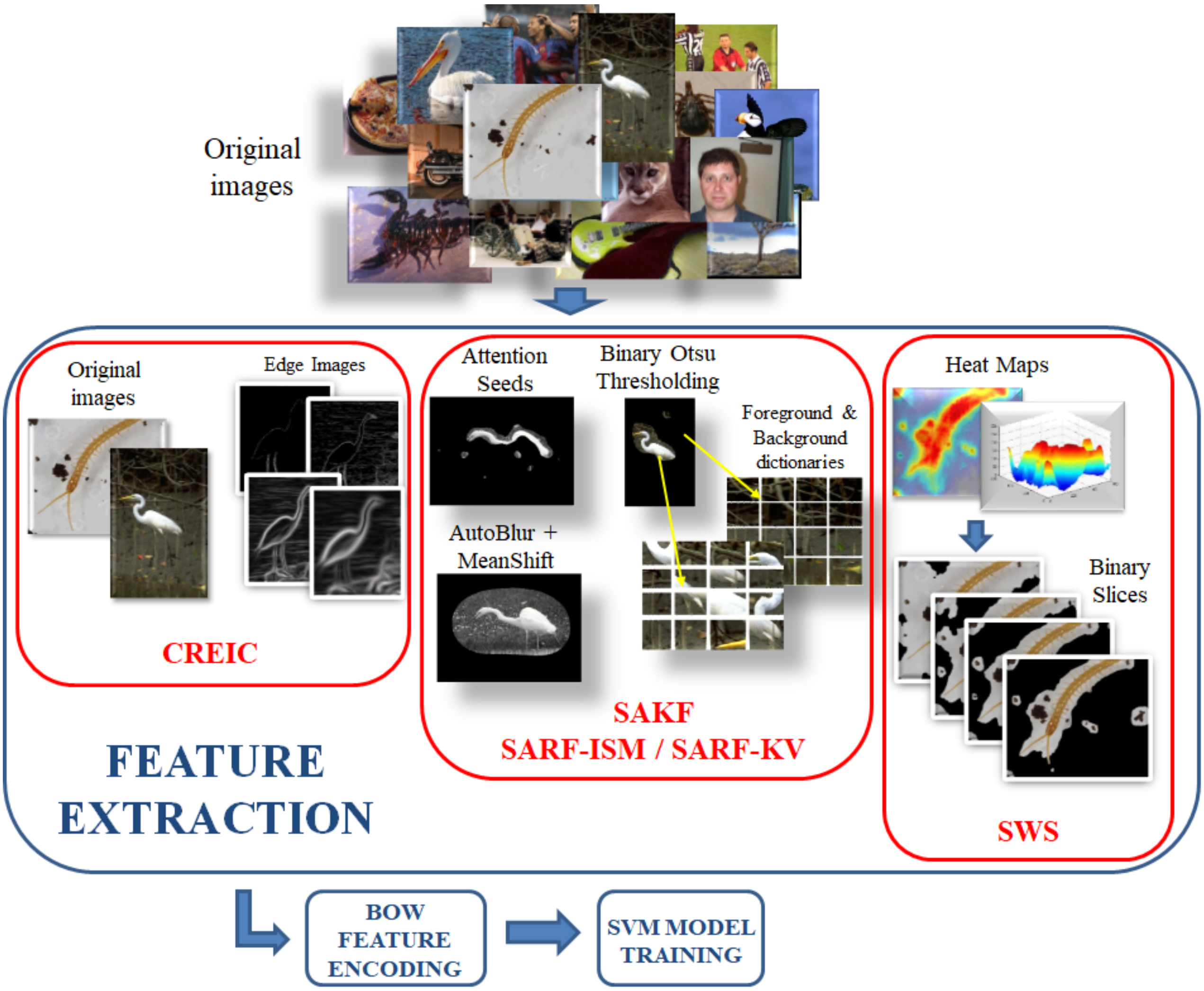

Figure 1: Overview of the contributions obtained at Feature Extraction level

2 Methodology and Contributions

2.1 CREIC: Compass Radius Estimation for Image Classification

Through the use of Support Vector Machine (SVM) classification, in this work we have demonstrated that the information extracted from the edges of the images using a compass operator is strongly dependent on the threshold chosen to obtain them, i.e. radius. Then, we assessed the performance that would have been achieved if the best radius parameter could be selected for each image of a dataset. Finally, we proposed *CREIC*, an original method to estimate the more suitable radius for a specific dataset. The method guarantees a better accuracy than the one obtained when using the radius suggested in the literature and also introduces savings in time when trying to find the right radius parameter. To retrieve more details of this contribution, the reader is addressed to [1], also available at [2] and to the papers [3] [4].

2.2 SAKF and SARF: Semantic Attention Filtering at Keypoint and Region level

If the information extracted from an image is mainly obtained from the foreground, the features used will yield a better classifier. Initially, we demonstrated how the blurring factor in the saliency map introduced by Xiaodi et al. [5], affects the image classification when it is evaluated with different blurring factors to separate the foreground and background information. Next, to avoid the saliency map dependency on blurring factor, we proposed *AutoBlur*, a method that automatically selects it for each image trying to retain the main features of the object of interest.

At this stage, we proposed three methods to filter the foreground key points. At pixel level, we introduced Semantic Attention Keypoint Filtering (*SAKF*), which filters the descriptors based on the distances from the foreground key points to both foreground and background previously created dictionaries. At region level, we proposed Semantic Attention Region Filtering (*SARF*) and we evaluated two different variants of it. The first one, *SARF* based on Intersection of Saliency Maps (*SARF-ISM*) is based on using the binary saliency map proposed by Xiaodi et al. using a blurring factor of 0.02 (hereafter, "Attention Seeds") combined with the regions obtained by segmenting the image with the *Mean-Shift* algorithm. Each of such regions sharing a common area with the attention seeds is considered to belong to the region of interest; otherwise, it is considered as part of the background. The second version, SARF based on Keypoint Voting (*SARF-KV*) determines if a region belongs to the background or foreground employing a voting system based on the distance from the foreground descriptors region to a foreground or a background dictionary. Both methods guarantee better accuracy than the baseline one, and they do not require optimising the saliency map parameters.

2.3 SWS: Semantic Weighted Schema

Considering a Saliency Map as a topographic map, we proposed a Semantic Weighted Schema (*SWS*) where we demonstrated that the information contained at different levels into a saliency map, named saliency information slices (SIS), can improve the final image classification results. The test has been performed on two saliency maps after evaluating different approaches on two subsets of ImageNet dataset [6]. For more details about this, the reader is addressed to [1] [2]. Combining the foreground information of different SIS obtained from the same or different saliency maps, it is guaranteed a higher accuracy than the one obtained for each SIS.

Apart from demonstrating that the accuracy is affected when information from SIS of different saliency maps is combined, it has been observed that the accuracy is halfway than when the SIS of each saliency map is used separately.

To finalise, we demonstrated that the number of information slices taken from the saliency map affects the dictionary optimisation, but that using more SIS does not imply better results in the final classification.

3 Results

The techniques presented in this work are recommended when the number of data available for the training is not high: the Bag of Visual Words framework is one of the best alternatives to the models obtained with Deep Learning techniques.

In this work, we compared the results of each contribution with the baseline method, which consisted on the accuracy obtained where only SIFT descriptors are extracted from all the image without any of the methods presented. The descriptors derived from a training set of images with the aid of the previous methods were hard-coded into feature vectors (through BoVW) that were used to train an SVM model with both Lineal and Intersection Kernels. Finally, the corresponding image test set is evaluated with the generated model, and it is obtained an accuracy value which is compared against the baseline one.

To retrieve more details about the results of each contribution, we addressed the reader to [1], also available at [2].

4 Conclusions and Future Work

Working in the Bag of Visual Words (BoVW) framework, in this paper we have presented three contributions focused in the feature extraction stage. We first illustrated how the information extracted from edge images affects to the image classification process. Then we demonstrated that the best parameter to compute the edge images, i.e. radius, is not the one recommended by the literature and we empirically calculated what it would be the ideal accuracy when ideal radius is selected for each image. Lastly, we proposed Compass Radius Estimation for Image Classification *CREIC*, which estimates a radius for calculating the dataset edge images, ensuring a better accuracy than the one obtained with the literature radius.

By introducing the Saliency Maps in the BoVW framework, we proposed four strategies to filter the foreground descriptors extracted from the image. We first proposed Semantic Attention Keypoint Filtering (*SAKF*), which filter the descriptors at the pixel level and then two variants of Semantic Attention Region Filtering (*SARF*). The first one, *SARF-ISM*, is based on Intersection of Saliency Maps and the second, (*SARF-KV*), is based on Keypoint Voting after computing foreground and background descriptors from regions segmented with Mean-Shift. Finally, considering a Saliency Map as a topographic map, we extracted the information contained at different levels, i.e. slices, and proposed a Semantic Weighted Schema (*SWS*). We evaluated the influence of the information obtained at different slices of the topographic map regarding their number and combinations. The results obtained outperform the baseline ones, demonstrating their validity to apply them at the feature extraction stage in a BoVW image classification process.

References

[1] E. Fidalgo Fernández, (2016). "Selección de información significativa para mejorar la clasificación de imágenes utilizando técnicas de Bag of Visual Words". ISBN: 978-84-9773-859-0, D.L.: LE-483-2016, p.205.

[2] E. Fidalgo Fernández, 2015. "Selection of relevant information to improve Image Classification using Bag of Visual Words", URL: *https://www.educacion.gob.es/teseo/mostrarRef.do?ref=1190235*

[3] E. Fidalgo, E. Alegre, V. González-Castro, L. Fernández-Robles, "Compass radius estimation for improved image classification using Edge-SIFT", *Neurocomputing*, Volume 197, 12 July 2016, pp. 119-135, ISSN 0925-2312, *http://dx.doi.org/10.1016/j.neucom.2016.02.045*

[4] E. Fidalgo, E. Alegre, V. González-Castro, L. Fernández-Robles, "Illegal activity categorisation in DarkNet based on image classification using CREIC method", *International Joint Conference SOCO'17-CISIS'17-ICEUTE'17, León, Spain, September 6-8, 2017, Proceeding pp. 600-609*

[5] H. Xiaodi; J. Harel, C. Koch, "Image Signature: Highlighting Sparse Salient Regions", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.34, no.1, pp.194,201, Jan. 2012, *doi:10.1109/TPAMI.2011.146*

[6] O. Russakovsky*, J. Deng* et al, (* = equal contribution), "ImageNet Large Scale Visual Recognition Challenge", *IJCV*, 2015