Performance Evaluation of Image Filtering for Classification and Retrieval

Falk Schubert[1] and Krystian Mikolajczyk[2]
[1]*EADS Innovation Works, Ottobrunn, Germany*
[2]*University of Surrey, Guildford, U.K.*

Keywords: Image Processing, Filtering, Enhancement, Logo Retrieval, Scene Classification.

Abstract: Much research effort in the literature is focused on improving feature extraction methods to boost the performance in various computer vision applications. This is mostly achieved by tailoring feature extraction methods to specific tasks. For instance, for the task of object detection often new features are designed that are even more robust to natural variations of a certain object class and yet discriminative enough to achieve high precision. This focus led to a vast amount of different feature extraction methods with more or less consistent performance across different applications. Instead of fine-tuning or re-designing new features to further increase performance we want to motivate the use of image filters for pre-processing. We therefore present a performance evaluation of numerous existing image enhancement techniques which help to increase performance of already well-known feature extraction methods. We investigate the impact of such image enhancement or filtering techniques on two state-of-the-art image classification and retrieval approaches. For classification we evaluate using a standard Pascal VOC dataset. For retrieval we provide a new challenging dataset. We find that gradient-based interest-point detectors and descriptors such as SIFT or HOG can benefit from enhancement methods and lead to improved performance.

1 INTRODUCTION

Significant progress has been made in image recognition and retrieval over past decades due to intensive studies of feature extraction methods, image representation and machine learning techniques. A number of alternative solutions have been proposed for each of the well established steps of the recognition and retrieval approaches. However, little research has been carried out on the quantitative influence of pre-processing steps which alter the image before applying the commonly used feature extractors in the computer vision applications mentioned above. Previous works include only basic filtering methods (e.g. blurring) employed in the context of very specific tasks such as face recognition (Heseltine et al., 2002; Gross and Brajovic, 2003; Kumar et al., 2011) or character recognition (Huang et al., 2007). Some open-source implementations of feature extractors also apply blurring as an initial step, but such pre-processing steps are never discussed in terms of quantitative performance gain in the respective papers. Besides these simple filtering techniques, there exists however a wide variety of more advanced image filtering techniques (e.g. bilateral filtering, cartoon-style or image-based rendering) in the domain of computer graphics which are not commonly used. Such filters, e.g. abstraction filters, have a direct impact on the image gradients which leads to a normalization of gradient-based descriptors. We therefore want to motivate the use of such advanced pre-processing filters in order to further increase the performance of computer vision applications, instead of re-designing or fine-tuning features for a specific computer vision task.

To better understand the quantitative difference image filtering techniques can generally make on the performance of feature extractors, we present in this paper a performance evaluation of a number of image enhancement or modification techniques applied to two common computer vision applications: scene recognition and logo retrieval. To our knowledge this is the first quantitative evaluation of such image filtering for pre-processing. Because image filtering is a data-driven or pixelwise local process, it is to be expected that the influence of image filtering also depends on the image content. We therefore evaluate using different datasets consisting of images of various categories. For scene recognition we evaluate using

Schubert F. and Mikolajczyk K. (2013).
Performance Evaluation of Image Filtering for Classification and Retrieval.
In *Proceedings of the 2nd International Conference on Pattern Recognition Applications and Methods*, pages 485-491
DOI: 10.5220/0004333104850491

the well-known Pascal VOC 2007 dataset (Everingham et al., 2010) which contains 20 different types of image scenes (e.g. natural scenes, man-made objects, etc.). For logo retrieval we evaluate using a dataset of 30 different logo classes (e.g. Volkswagen, BMW, Coca Cola, etc.) which consists of real images of these logos captured in normal life (i.e. the images were taken from personal and professional photographs downloaded from Flickr).

The paper is structured as follows: In section 2 we briefly discuss the different filtering techniques which we will consider. In section 3 we give details about the implementation of the two computer vision applications. In section 4 we discuss the evaluation protocol and discuss the results on the benchmark datasets.

2 FILTERING TECHNIQUES

The type of normalization and hence the effect on subsequent steps of the recognition system, in particular feature extraction, strongly depends on the way the filter modifies the image. We focus on three categories of filters: boosting gradients, suppressing gradients and enhancing color. These types are motivated by the fact that the most successful features in computer recognition applications are based on gradients (Everingham et al., 2010) (e.g. Harris, Hessian, HOG, SIFT, DAISY) and color (e.g. Color-SIFT). For each of the filter types we consider different methods which we discuss in the following.

2.1 Boosting Gradients

In the computation of HOG or SIFT the strength of the gradient is used to weight the corresponding bins in the histograms. Hence boosting important gradients can increase their importance in the image descriptors. We consider two different variations to increase the strength of gradients: convolution with sharpening kernels and tonemapping (Fattal et al., 2002), which is based on a compression, where weak edges are boosted and strong ones are reduced. The first is a fairly simple and well-known filter. The second one is a more complex filtering technique which roughly works as follows. A new gradient field is computed from the existing gradients $H(x,y)$ according to the formula (Fattal et al., 2002) :

$$G(x,y) = \nabla H(x,y) \cdot \varphi(x,y)$$

The attenuation factor φ modifies the existing gradients $H_k(x,y)$ at different resolution scales k of an image and is computed as (Fattal et al., 2002) :

$$\varphi_k(x,y) = \frac{\alpha}{\|\nabla H_k(x,y)\|} \left(\frac{\|\nabla H_k(x,y)\|}{\alpha} \right)^{\beta}$$

From this gradient field G an image is reconstructed using the Poisson equation:

$$\nabla^2 I = \operatorname{div} G$$

The parameters α and β control the amount of attenuation of large gradients and magnification of small ones. The effect of tonemapping (tmo) is visualized in the right column of Fig. 2.

2.2 Suppressing Gradients

Eliminating only weak gradients results in smooth image segments and cartoon-like stylization. On such images feature detectors generate interest points mainly on dominant image structure. These interest points tend to be more stable under different variations (e.g. pose variations). This effect can help to focus a learning process on the important image structures, leading to a better visual recognition despite the loss of information. We consider four different gradient suppressing filters: Gaussian blurring, median filtering, bilateral filtering (Tomasi and Manduchi, 1998) and weighted-least-squares filtering (wls) (Farbman et al., 2008). The first three are often used as pre-processing filters. However, the impact of these preprocessing steps are rarely discussed or evaluated. The fourth filter is an advanced edge-preserving filter superior to the standard bilateral filtering technique. The image is obtained via an iterative optimization procedure (i.e. weighted least squares) which minimizes the following cost function C (Farbman et al., 2008):

$$C = \sum_{(x,y)} \Bigg([I(x,y) - O(x,y)]^2 + \lambda \left[u_O(x,y) \left(\frac{\partial I}{\partial x} \right)^2 + v_O(x,y) \left(\frac{\partial I}{\partial y} \right)^2 \right] \Bigg)$$

The first term of the cost function ensures that the resulting image I is visually similar to the original input image O. The second term acts as a regularizer which suppresses gradients along x- and y-direction in the resulting image at all locations where the input image O contains weak gradients. These locations are controlled by the spatially varying weights u and v:

$$u_O(x,y) = \left(\left| \frac{\partial O(x,y)}{\partial x} \right|^{\alpha} + \varepsilon \right)^{-1}$$

The formulation for v is analogous to u, just the partial derivative is along y direction. The manually chosen parameter α controls the strength of the smoothing effect. The constant ε prevents division by zero.

All other locations which contain dominant gradients are left unchanged. The parameters λ and the weight functions u and v control the amount of smoothing. The effect of this edge-preserving or weighted least squares (wls) filtering is visualized in the center column of Fig. 2.

To better understand the difference of gradient suppression and gradient enhancement an illustration is given in Fig. 1. Intensity values along line scans of an example image are shown. The left plot shows the original intensity values, the center one shows the intensity values after tonemapping, the right one shows the intensity values after applying the WLS filter. Filters such as tonemapping boost weak gradients and keep the dominant ones unchanged. The filters suppressing gradients such as abstraction filters keep dominant gradients but significantly smoothen small gradients. The choice of the influence of boosting and suppression is clearly arbitrary and depends on the image content and the subjective taste of the user. In our experiments in sec. 4 we selected these parameters manually prior to all experiments without focussing on increasing the performance but purely on visual appearance (e.g. the settings of the WLS filter where chosen to clearly suppress weak gradients whereas the tonemapping set to clearly boost them). In Fig. 2 examples of the filtered images are depicted for each dataset. In future work it would be very helpful to have a learning-based process that find these settings automatically. However, the contribution of this paper is a quantitative performance evaluation of the impact of the filtering techniques independent whether they have been chosen manually or automatically.

2.3 Enhancing Colors

Using color in image descriptors was reported to significantly improve the recognition results (Yan et al., 2012). Pictures are often taken with sub-optimal color settings due to simplistic auto-exposure controls and auto-white-balancing. A post-processing step can help recover or improve the contrast of the image if all details and structures are captured without saturation. A histogram normalization step (which we call colorboost) can be used to equalize the colors within an image and bring out much better detail. We use the method described in (Horváth, 2011), which is very robust and parameter free. The filtered color images also show better contrast when converting them to grayscale. As all images in our benchmark datasets are colored, we can therefore evaluate this color-normalization also for descriptors which only use grayscale images.

3 APPLICATIONS

A straightforward approach to investigate the impact of image filtering techniques, could either consist of a simple toy application (e.g. simple nearest neighbor search within a pool of features) or some heuristic measurements on the feature vector (e.g. variations of individual feature dimensions or intra/inter-class variance). However, in our opinion conclusions drawn from such experiments cannot really be generalized to other realistic applications. We therefore propose to evaluate the impact of image enhancement on the performance of typical computer vision tasks. We considered two different applications: scene recognition and image retrieval. Both applications share a search task or matching step based on features that are computed. In the first case this matching step is based on a learning process, whereas in the second case a simple distance measure is used. In the following a brief summary of the implementation of each application is given.

3.1 Image Retrieval

We employ a bag-of-words representation (Sivic and Zisserman, 2003) which has become the state-of-the-art for fast scalable retrieval and classification tasks. The different computation steps can be summarized as follows:

1. detect interest points (Hessian and Harris-Laplace)
2. extract SIFT features at the interest points
3. generate visual words from the image features of the whole training dataset using randomly selected clusters
4. compute an inverted file index from histograms of visual word occurrences for every image
5. search with new image as query using L2-norm on the index signatures

Many powerful extensions have been proposed in the past to enforce geometric consistency or expand queries (Philbin, 2010). However, the baseline approach as described in (Sivic and Zisserman, 2003) is sufficient to demonstrate the benefit of using preprocessing filters.

3.2 Scene Classification

In image classification experiments we use the approach from (Yan et al., 2012) that has proven very successful in various classification benchmarks. The different computation steps can be summarized as follows:

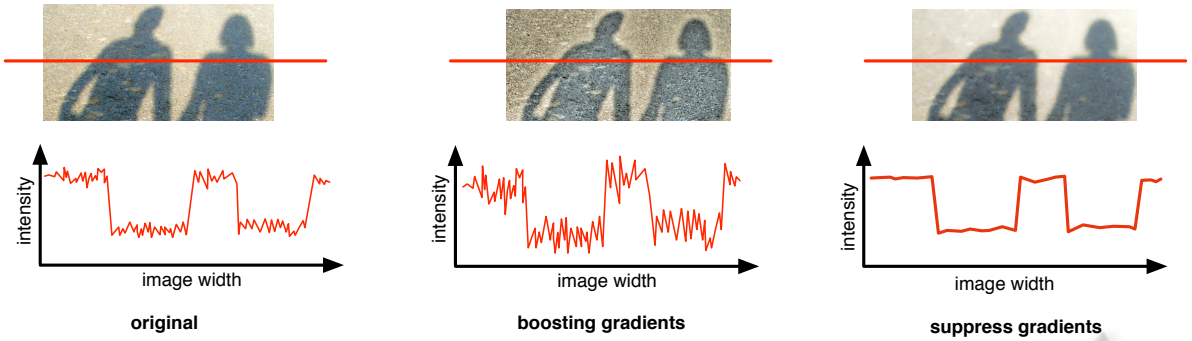

Figure 1: Intensity values (diagrams on bottom) along the line scans across the image (red) are shown for different filters: original (left), boosting (center) and suppression (right) of gradients.

1. compute local image descriptors (e.g. SIFT, CSIFT, etc.) on a uniform dense grid
2. generate visual words from the image features of the whole training dataset using randomly selected clusters
3. compute a histogram of visual word occurrences for every image
4. compute spatial pyramid match kernel (Lazebnik et al., 2006) from the histrograms
5. train SVM with χ^2 kernels
6. classify new image using combination of multiple kernels

4 EVALUATION

In this section we evaluate the impact of the image filtering techniques on the recognition applications. For classification and retrieval different benchmarks have been established in the literature. We evaluate the impact of the image filtering techniques using standard evaluation protocol of the respective datasets used for the two applications. For the scene recognition the Pascal datasets are considered as the gold standard. We use the Pascal VOC 2007 dataset (Everingham et al., 2010) which contains 20 different types of image scenes (e.g. natural scenes, man-made objects, etc.). These variations can be considered challenging enough to allow drawing valid conclusions about the impact of the filtering techniques.

For the image retrieval task, we use an own dataset for many reasons. Typical retrieval datasets (e.g. the Oxford Building dataset (Philbin, 2010) or the Flickr1M dataset as used in (Jégou et al., 2008)) either address the scalability of the retrieval task or they are designed for the retrieval of a very specific image. In the first case this means that they are very big (e.g. Flick1M dataset with 1 million images (Jégou et al., 2008)) and the goal is to show that the retrieval engine is capable of finding many images that are similar to the input image. In the second case these datasets contain images of specific objects (e.g. buildings like in the Oxford building dataset (Philbin, 2010)) which might have been taken from different view points and the goal is to find all instances of the object shown on the input image.

We would like to generalize the retrieval task further by allowing more variation to the retrieved objects but still ensure that the look of the object is well defined. This is the case in logo retrieval. The logos are like objects (e.g. building) but they can have different appearances (e.g. a painted logo or printed logo) and yet belong to the same logo label. There exist a few datasets for logo retrieval however some of which are too simple (e.g. only contain synthetic images (Jain and Doermann, 2012)) or which are not consistant (e.g. logos have the same label where the logo changed its design over time (Kalantidis et al., 2011)). We therefore provide a new logo retrieval dataset with 30 different logo classes which has roughly the same size or variation as the existing datasets (e.g. the Flick27 logo dataset with 27 logos classes (Kalantidis et al., 2011)).

4.1 Scene Classification

Improvements in scene classification are evaluated using the evaluation protocol from Pascal VOC 2007 dataset (Everingham et al., 2010). More specifically the “average-precision” (AP) which is the area under the precision-recall curve (Everingham et al., 2010) is computed for each scene class for both the original, unaltered images and for all filtered ones. In this experiment we considered four different filters (blur, colorboost, bilateral, wls). The filteres were applied to each training and test image and evaluated separately with constant settings for all experiments. The results are summarized in Tab. 1.

Table 1: Comparison of recognition performance (AP) on VOC 2007 using best performing filter and original images. In the fourth column the difference of AP between filtered and original images are given.

class	filter	original	diff	filter name
aeroplane	64.9	64.4	+0.5	bilateral
bicycle	56.2	52.9	+4	wls
bird	43	37	+6	bilateral
boat	55.5	52.5	+3	colorboost
bottle	19	14.3	+4.7	bilateral
bus	43.4	43.1	+0.3	colorboost
car	69.4	68	+1.4	bilateral
cat	45.4	46.4	-1	colorboost
chair	42.4	41.6	+0.8	bilateral
cow	23.9	21.8	+2	wls
table	31.9	29.5	+2.4	bilateral
dog	35.8	36.1	-0.3	colorboost
horse	64.6	65.2	-0.7	colorboost
motorbike	52.6	49	+3.6	wls
person	78.7	77.8	+0.9	bilateral
plant	22.6	18.6	+4	bilateral
sheep	26.6	28	-1.4	bilateral
sofa	33.7	32.6	+1.1	blur
train	64.3	63.2	+1.1	bilateral
tv	39.9	39.2	+0.7	colorboost

For 16 out of 20 classes in Tab. 1 filtered images produce better results than the original ones. Gradient suppression (e.g. bilateral or wls filters) in particular improves the AP performance by up to 6%. This can be explained by the elimination of weak, noisy gradients using abstraction filters such as bilateral filtering. For instance, many of the images in the class "bird" were captured with background such as vegetation and nature, which contain many fine detailed gradients that are irrelevant for the classification. Focusing the descriptors on dominant gradients (e.g. stems from trees and not the leaves, bird shape and not the feathers) helps to discriminate these images. Again we note that the an automatic choice of the best performing filter would be required for practical applications. However, in this experiment we are more interested on the quantitative performance differences, which indicate how much mAP can be gained by a good choice of image filtering for preprocessing. The filter parameters were manually chosen prior to all experiments without focussing on increasing the performance but purely on visual appearance to achieve clearly visible filtering effects.

4.2 Image Retrieval

For reasons mentioned above, we collect our own benchmark dataset with images that present particular challenge to the descriptors due to various rendering methods (e.g. logo is painted on a wall or carved out of metal) which introduces more appearance variations (see Fig. 2 and Fig. 1 for some examples). In such cases, image filtering is especially expected to aid the matching process. The dataset consists of 30 random logos classes from well known brands (e.g. Coca Cola). For each logo 10 random images were pooled out of 1000 images downloaded from www.flickr.com using the logo name as the search query. For all 300 images of the dataset the occurences of the logos are labeled. The retrieval task is to use each labeled logo and retrieve all the other ones with the same label. We use the same protocol for the generation of the index and evaluation of the retrieval performance as in (Sivic and Zisserman, 2003). Similarly to the evaluation of scene classification all filter settings were constant for all images and were chosen prior to running the experiments. In Tab. 2 the summarized mean-average-precision (mAP) values are listed separately for the two interest point detectors (Harris-Laplace and Hessian-Laplace) used in the experiment. For each query image an AP value (Everingham et al., 2010) is generated which is then averaged (mAP value) across all queries belonging to the same logo label. We further average these mAP values over all logo labels to generate a single score for each filter. We can observe that gradient suppression filters, in particular median and wls, improve the retrieval by up to 8%. The performance gain depends on the type of interest point detector, but the general tendency is the same. It is important to note, that the overall performance of $mAP \approx 45\%$ is not very high compared to systems with geometric verification or query expansion (Philbin, 2010). However, in this experiment we are interested in relative performance differences between filtered and unaltered images. Although the overall performance across a collection of 30 very different logos consistently improves by using wls filtering, we noticed that certain logo types benefit more than the others. Car logos (e.g. Porsche) which do not vary as much in their rendering form (e.g. car logos are usually printed on badges and not other material like T-Shirts) improve by 58.1% (mAP for "Porsche" logo using original images is 36.2% and 94.3% using wls filtering).

5 CONCLUSIONS

The results from the evaluation indicate that image filtering significantly improves the matching and classification performance. Furthermore the amount of improvement and the type of best performing filter

Figure 2: Sample images (original and filtered) from the logo dataset (top 2 rows) and Pascal VOC 2007 (bottom 2 rows).

Table 2: Mean-Average-Precision (mAP) listed for each filter and interest point detector. Behind each mAP score, the difference to the original (top row) is given.

filter name	Harris (diff)	Hessian (diff)
original	32.4	38.4
bilateral	35.5 (+2.9)	39.5 (+1.1)
blur	33.7 (+1.3)	39.8 (+1.4)
colorboost	33.3 (+0.9)	41.4 (+3.0)
median	35.4 (+3.0)	44.2 (+5.8)
sharpen	29.7 (-2.7)	36.1 (-2.3)
tonemapping	31.9 (-0.5)	39.4 (+1.0)
wls	40.4 (+8.0)	46.9 (+8.5)

depends on the image category (e.g. natural scenes, synthetic images). For the recognition for each class different filters perform best. For the retrieval certain types of logos benefit more from filtering than others. In future work we would like to further investigate the impact of image filtering on different types of interest point detectors and features. Also we would like to develop an automatic selection process which finds the best suiting filter type and parameter settings given a training dataset. Last but not least, we would like to include other and notably larger datasets for the retrieval and consider other applications like object detection.

REFERENCES

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *IJCV*.

Farbman, Z., Fattal, R., Lischinski, D., and Szeliski, R. (2008). Edge-preserving decompositions for multi-scale tone and detail manipulation. *SIGGRAPH*.

Fattal, R., Lischinski, D., and Werman, M. (2002). Gradient domain high dynamic range compression. In *SIGGRAPH*.

Gross, R. and Brajovic, V. (2003). An image preprocessing algorithm for illumination invariant face recognition. In *AVBPA*.

Heseltine, T., Pears, N. E., and Austin, J. (2002). Evaluation of image pre-processing techniques for eigenface based face recognition. *ICIG*.

Horváth, A. (2011). Aaphoto. *http://log69.com/aaphoto_en.html* .

Huang, B. Q., Zhang, Y. B., and Kechadi, M. T. (2007). Preprocessing techniques for online handwriting recognition. In *ISDA*.

Jain, R. and Doermann, D. (2012). Logo retrieval in document images. *Document Analysis Systems, IAPR International Workshop on*, 0:135–139.

Jégou, H., Douze, M., and Schmid, C. (2008). Hamming embedding and weak geometric consistency for large scale image search. In *ECCV*.

Kalantidis, Y., Pueyo, L., and Trevisiol, M. (2011). Scalable triangulation-based logo recognition. *ICMR*.

Kumar, M., Murthy, P., and Kumar, P. (2011). Performance evaluation of different image filtering algorithms using image quality assessment. *IJCA*.

Lazebnik, S., Schmid, C., and Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scenes and categories. *CVPR*.

Philbin, J. (2010). *Scalable Object Retrieval in Very Large Image Collections*. PhD thesis, University of Oxford.

Sivic, J. and Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. In *ICCV*.

Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *ICCV*.

Yan, F., Kittler, J., Mikolajczyk, K., and Tahir, A. (2012). Non-sparse multiple kernel fisher discriminant analysis. *JMLR*.