ADAPTIVE DENOISING FILTERING FOR OBJECT DETECTION APPLICATIONS

Simone Milani, Riccardo Bernardini, Roberto Rinaldo∗

DIEGM - University of Udine
Via Delle Scienze, 208 - 33100 Udine - Italy
e-mail: `simone.milani@uniud.it, bernardini@uniud.it, rinaldo@uniud.it`

ABSTRACT

The widespread of augmented reality applications, cognitive video surveillance, autonomous or supportive navigation systems, has increased the importance of accurate object detection algorithms. However, the presence of noise depending on the characteristics of the acquisition device, on lighting intensity and directions, and on weather conditions, could severely degrade the performance of such applications. As a matter of fact, effective ad-hoc denoising strategies are required since traditional noise removal algorithms designed to improve the quality of the image, could even worsen the accuracy of detection.

This paper presents a low-cost adaptive filtering strategy that adapts the characteristics of the filter depending on the impact of each image region on the feature sets. This solution permits improving the correct detection percentage of approximately 30% with respect to using noisy images. The approach is generally intended for object detection algorithms based on Histogram-of-Oriented-Gradients (HOG) and can run in real time on a limited complexity hardware.

Index Terms— denoising, object detection, adaptive filtering, saliency map, HOG

1. INTRODUCTION

Object detection is nowadays a key element in many multimedia applications such as augmented reality, video surveillance, autonomous or supportive navigation systems, and data classification (just to mention some of them). In these applications, the acquired scene is analyzed identifying the present objects, their positions in the space, and their relations or actions.

However, the efficiency of such systems can be severely compromised by the presence of noise and by the overall quality of the captured image. Augmented reality applications are usually intended for mobile devices which may be enabled with a low-cost camera (with high-levels of sensor noise) and usually acquire the visual information in uncontrolled environments (limited or excessive illumination, adverse weather conditions). As for autonomous or supportive navigation systems, they need to perform very effectively in adverse conditions for safety reasons. Fig. 1 shows the effects of Gaussian noise on the object detection algorithm in [1]; it is possible to notice that the additional noise prevents the algorithm from identifying all the elements belonging to the class `persons` which are present in the scene. As a matter of fact, it is necessary to design efficient processing algorithms to reduce the amount of noise in the image.

∗This work was partially supported by the POR FESR 2007 – 2013, Friuli Venezia Giulia Regional Project "Barcotica."

(a) (b)

Fig. 1. Performance of object detection under noisy conditions. The adopted query model is `persons`. a) result without noise b) result with additional Gaussian noise (SNR=-18.06 dB).

Traditional denoising strategies are based on the assumption that images present a low-pass characteristic while noise covers a wider spectrum of frequencies. As a matter of fact, it is possible to reduce the energy of the noise by processing the image data via a low-pass filter that discards the frequency band that is occupied by the noise signal only. Better results can be obtained by building a matched filter once the filtering strategy has estimated a model for the original signal. However, this filtering operations could alter the features computed on the image; as a result, objects could not be detected any more or detection errors could become more frequent. Many object detection methods are based on Histograms of Oriented Gradients (HOG) [2]. Unfortunately, low-pass filtering strongly alters the statistics of gradients and, as a result, the generated features could not match the features stored in the object models.

It is possible to overcome this problem by designing an ad-hoc filtering strategy that takes into consideration which data are the most important in the computation of object features. The current paper presents a low-complexity denoising strategy designed to enhance the performance of object detection. This approach has been intended for autonomous or supporting navigation systems (where execution time is an important issue) or for mobile augmented reality applications (where low complexity increases the battery life and, as a consequence, the autonomy of the device). The proposed approach identifies those parts of the image signal that generate the most crucial descriptors. Then, the denoising filter is adjusted in order to remove most of the noise whenever no harm can be done to object descriptors and to preserve most of the details whenever important information is present. The approach has been validated using the object detection algorithm in [1], although further validation is currently ongoing using different object detection algorithms. Experimental data shows that the denoising strategy can improve the correct detection performance of the algorithm up to 30% at high noise levels.

In the following, Section 2 overviews some of the object detection strategies that have been proposed in the literature. Section 3

978-1-4673-2533-2/12/$26.00 ©2012 IEEE

describes the proposed algorithm, whose test results are reported in Section 4. Conclusions (Section 5) end the paper.

2. OBJECT DETECTION ALGORITHMS

Object detection has been studied for about four decades producing a wide range of object detection techniques. This research effort has been fostered by the many application scenarios where object detection can be employed (e.g., active video surveillance, assisted or autonomous drive, database search, data classification) and by the need of increasing the robustness of existing algorithms to different lighting conditions, poses, scales, locations, and geometries (deformable objects). As a matter of fact, a modern object detection algorithm can be divided into five parts: pre-processing and normalization, local rectification and compensation of small shape variations, computation of descriptor set, machine learning classification, and post-processing to fuse multiple detections.

The pre-processing and rectification phases are very important since they permit compensating light changes and variations in the shape and positions of the object. Usually they include normalizing operations like gamma correction, local contrast normalization, and highlight suppression.

Then, the resulting pixels are processed to generate a set of descriptors for the objects in the image. Different descriptors have been proposed in the literature, like binary patterns from image pixels or the outputs of a set of steerable filters. The most recent algorithms employ edge orientation histograms. The input image is divided into cells, and for each pixel in the cell the algorithm computes the edge orientation from the image gradient. The histogram of these values is then computed, and the operations are iterated at different image scales. Histograms of gradients are nowadays widely adopted and provide the basis for the popular SIFT [3], HOG [2], and Generalized shape Context methods.

In the classification phase, the descriptors are classified partitioning the feature space into regions. In this case, Support Vector Machine (SVM) classifiers are widely adopted and perform quite well provided that an adequate learning phase is operated. Since the operation is iterated at different scales and multiple detections could emerge from the object detection process, it is necessary to combine and polish the responses of the classifier in order to make the detection more accurate.

All these elements play a significant role in the final accuracy, but in this paper we are going to focus on the first phase. In fact, noise and alterations might severely compromise the efficiency of the algorithm since they significantly alter the generated descriptors. As a result, the experimental data do not fit the partitioned space any longer, and several false positives and negatives can be found in the detection phase. Pre-filtering the image before object detection could be an interesting solution provided that the adopted filters take into consideration the following object detection operations.

A denoising method for biological macromolecule detection has been proposed in [4] where a set of rotation-equivariant nonlinear filters is employed to denoise contours and perform a rapid object detection in microscopical images. The approach proposed in [5] adopts a noise reduction strategy for pavement images based on wavelet packets.

In this work, we focus on detection schemes based on histogram of gradients because of their diffusion and the possibilities they offer in implementing real-time object detection systems. Since descriptors are computed from gradient orientation and color information, it is necessary to design edge-preserving denoising algorithms in order to prevent a decrease of the detection performance. In the following section, we will show how this is possible by designing an adaptive strategy based on simple filters.

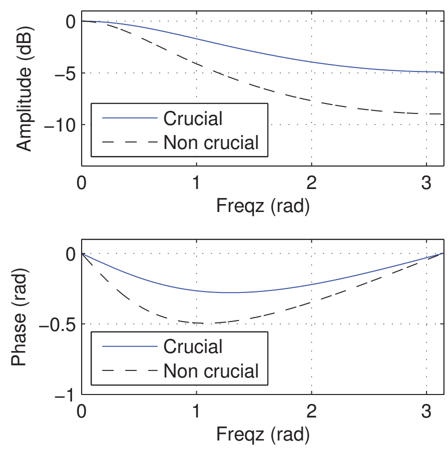

Fig. 2. Frequency response of the adopted filters.

3. THE PROPOSED DENOISING ALGORITHM

As previously mentioned, most of the object detection algorithms are based on analyzing the distribution of gradients and colors along object borders. As a matter of fact, an effective denoising system must take into consideration this fact by removing most of the noise from the scene while preserving the information that characterizes the features of the objects, i.e., pixel values along the edges.

A significant amount of edge-preserving filters can be found in literature (e.g., bilateral filtering in [6], just to mention one of the most famous). Other wavelet-based denoising filters have been considered as well.

In our approach, the need of implementing a limited complexity architecture led us to design a set of possible filters that are adaptively chosen. This strategy permits obtaining good performances in terms of accuracy in the object detection without requiring complex estimations of weights and parameters for the filter. The key issue in its design is the preservation of edge information and the avoidance of excessive smoothing whenever the contrast level between the foreground object to be detected and the background is low. As an example, let us consider Fig. 1, where the noise does not allow to identify the second person since it presents a lower contrast with the background. In many of these cases, keeping the noisy version of the image proves to be a better strategy since smoothing the signal with a filter would decrease the performance of the object detection algorithm.

In the proposed approach we adopted a simple separable 2D filter that combines two 1D single-pole IIR filters. The pole of the filter changes dynamically according to a saliency function $\mathbf{S}$ that partitions the input image into critical and non-critical blocks.

Critical blocks are characterized by strong edges that separate regions presenting a small color difference. Moreover, the regularity of edge directions proves to be an important factor to identify regions where an excessive filtering could severely degrade the object detection.

In order to assign a saliency value to the different regions of the image $I(x, y)$, the denoising strategy computes horizontal and vertical Sobelian gradients (named $S_x(x, y)$ and $S_y(x, y)$, respectively) on the whole image using a 3×3 Sobel operator. Then, these two gradient maps are merged into a common measure of the gradient strength named

$$G(x, y) = round\left(\frac{(|S_x(x, y)| + |S_y(x, y)|)}{16}\right) \quad (1)$$

where the rounding operation is used to remove the smallest gradients. Similarly, the orientation of borders $A(x, y)$ is also computed as $A(x, y) = \tan^{-1}(S_y(x, y)/S_x(x, y))$. These values are used to compute the regularity map of edges

$$R_A(x, y) = \frac{|A(x, y) - A(x + 1, y)| + |A(x, y) - A(x, y + 1)|}{2}. \quad (2)$$

The denoising algorithm then computes the 85-percentile T_g of the probability mass function of $G(x, y)$ for the current image. By selecting those pixel positions (x, y) such that $G(x, y) > T_g$, it is possible to consider only those positions that present relevant gradient information.

For every pixel position (x, y) such that $G(x, y)$ is greater than T_q, the algorithm evaluates the color differences along the edge orientation, i.e., it computes

$$\begin{aligned} C(x, y) = &\frac{|R(x + \delta_x, y + \delta_y) - R(x - \delta_x, y - \delta_y)|}{3} \\ &\frac{+|G(x + \delta_x, y + \delta_y) - G(x - \delta_x, y - \delta_y)|}{3} \\ &\frac{+|B(x + \delta_x, y + \delta_y) - B(x - \delta_x, y - \delta_y)|}{3} \end{aligned} \quad (3)$$

where $(\delta_x, \delta_y) \in \{-1, 0, 1\}^2$ is a displacement array aligned along the normal direction to the edge in (x, y). This parameter proves to be extremely important in identifying those objects that can not be easily detected from the background since a low value of $C(x, y)$ denotes a limited contrast around the edge.

The input noisy image is divided into 32×32 pixel blocks (indexed with the variable b), and for each of these blocks the algorithm computes the variance $\sigma_{R,b}$ of $R_A(x, y)$ in the current block. The algorithm also evaluates the parameter $\overline{C}_b$, which averages the values $C(x, y)$ for those pixel locations (x, y) in the b-th block such that $G(x, y) > T_g$.

From these variables, it is possible to build a saliency vector $\mathbf{S}_b = [\sigma_{R,b}, \overline{C}_b]$ that defines the characteristics of the image block and its impact in the object detection process. The vector space of $\mathbf{S}_b$ is then partitioned into regions that are associated to different labels and filter parameters. As a consequence, it is possible to write the filtering operation for the b-th image block $I_b(x, y)$ as

$$I_{f,b}(x, y) = h(x, y; \mathbf{S}_b) * I_b(x, y). \quad (4)$$

In this work, we partition the space into two subspaces defining two possible filters. In case $T_{r,m} < \sigma_{R,b} \leq T_{r,m}$ and $\overline{C}_b < T_C$, the block is labelled as crucial. Otherwise, the block is considered not crucial.

During the filtering process, each 32×32 block is extended in order to include some pixel positions for the transient phase of the filter. In order to obtain an extremely low computational complexity and avoid excessive cutoffs, we adopted two IIR filters with transfer function $H(z) = (1 - p_b)/(1 - p_b z^{-1})$, where the single pole is computed as

$$\begin{aligned} p_b &= p_0 + k \cdot \tfrac{3}{\rho} \quad \text{for crucial blocks and} \\ p_b &= p_1 + k \cdot \tfrac{3}{\rho} \quad \text{for non crucial blocks.} \end{aligned} \quad (5)$$

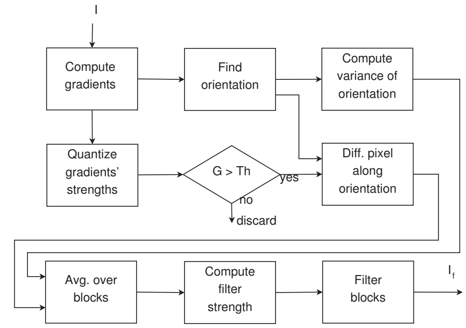

Fig. 3. Block diagram of the proposed algorithm.

The parameter $\rho \in [0, 1]$ is the average of the vertical and horizontal correlation coefficients for the pixels in the current block, i.e.,

$$\rho = .5\,\frac{E[I(x, y + 1) \cdot I(x, y)]}{\sigma_I^2} + .5\,\frac{E[I(x + 1, y) \cdot I(x, y)]}{\sigma_I^2}, \quad (6)$$

while $p_1 = 2\ p_0$ and p_0 depends on the variance of the additional noise (in our tests it was set to 0.2). Each color component is filtered independently. Fig. 2 reports the frequency response of the adopted filters with $\rho = 0.8$. It is possible to notice that for crucial blocks (i.e., blocks with limited orientation variance and small color gradients across edges) the adopted filter avoids smoothing color information too much (since high frequencies are preserved). For blocks with stronger color differences and orientation variance, the filter presents a stronger low-pass behavior.

The whole strategy is summarized in the block diagram reported in Fig. 3.

4. EXPERIMENTAL RESULTS

The proposed approach has been tested using the object detector in [1]. The adopted system represents objects using mixtures of deformable part models. More precisely, the whole algorithm inherits the key idea of the HOG-based approach [2] but extends object models introducing deformable parts. The HOG-based descriptors are then classified using a latent SVM classifier. Although tests were carried out using this approach only, we believe that the proposed solution can be extended with success to any HOG-based object detector. Different object classes were loaded (based on the PASCAL VOC 2006 dataset [7]) with the corresponding training and test sets. For each class, 10 different images were randomly selected and independent Gaussian noise was added with different SNR levels (10 different realizations were generated for each level). Detection accuracy was measured computing the ratio between the correctly identified object with respect to the real objects in the scene, i.e.,

$$P = \frac{\text{number of correctly-detected objects}}{\text{number of objects from ground truth}}. \quad (7)$$

More precisely, since the object detection algorithm identifies a region in the image which is supposed to include the searched object, in our tests we assumed that identification was correct when the detected region matched at least 80% of the corresponding region in the ground truth image. The performance reported in the following was obtained averaging experimental results for the different realizations.

Figure 4 reports the percentage of correct identifications and localizations for the object classes `person`, `car`, `horse`, and `cow`.

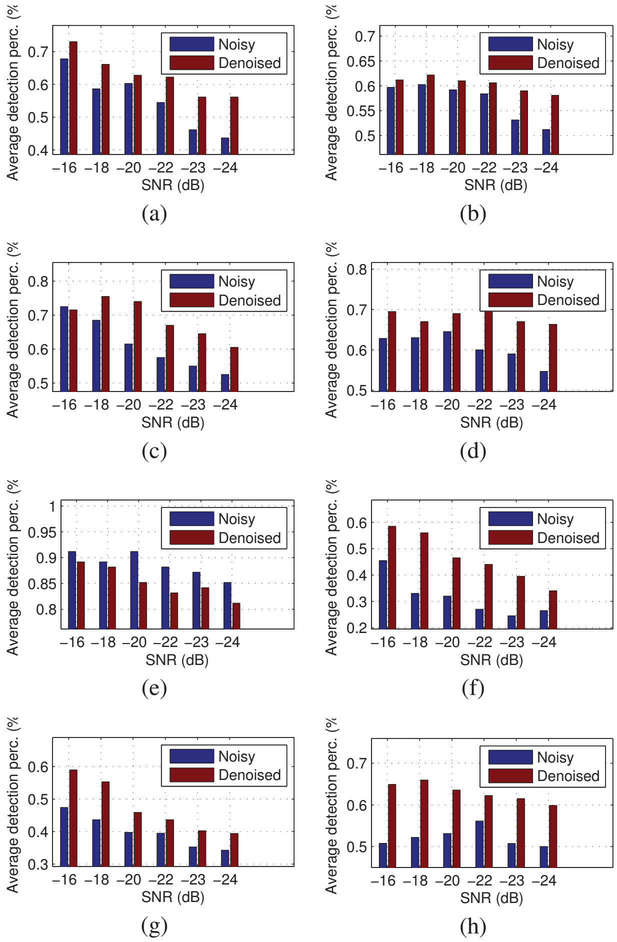

Fig. 4. Average percentage P of correctly-detected objects under diverse noise conditions for noisy and denoised images. a) `person` b) `car` c) `bus` d) `motorbike` e) `bicycle` f) `horse` g) `cow` h) `sheep`.

It is possible to notice that a significant performance improvement can be obtained at low SNR level. For the dataset `horse`, accuracy has a relative 60% improvement for an SNR decrement -23 dB (see Fig. 4). In the case of dataset `car`, the improvements are smaller since the absolute accuracy decreases less significantly with the decrement of the SNR level. Similar improvements can be noticed for the classes `person` and `cow`, where we found relative precision increments up to 30%.

In order to provide a visual evidence, Fig. 4 reports the results of object detection on the noisy image of Fig. 1 after denoising. It is possible to notice that the proposed approach permits identifying the second person in the image. We also evaluated the performance of a generic wavelet-based image denoising applied on the image (see Fig. 4). It is possible to notice that no improvements are obtained with respect to the noisy version in Fig. 1.

As for the computational complexity, the filtering operations runs in 13 ms on a 3.4 GHz processor on an RGB image with PAL format. As a matter of fact, it supports denoising of video frames at approximately 60 Hz frame rate.

Fig. 5. Performance of different denoising algorithms for SNR=-18 dB. a) proposed b) generic wavelet-based denoising.

5. CONCLUSION

The paper presented a low-complexity denoising strategy aimed at improving the performance of object detection on images taken in an uncontrolled environment. The approach relies on an adaptive selection of filters depending on the local characteristics of the image. More precisely, the adopted filters preserve regions with small contrast from excessive smoothing. The approach performs quite well on HOG-based object detectors permitting sensible performance improvements at high noise levels. In the future, research activity will test the approach using different object detection algorithms and will increase the number of possible filters that can be adopted. Moreover, filter parameters will be optimized within a train-validate-test paradigm involving multiple loops of cross validation to maximize the effectiveness of the final algorithm.

6. REFERENCES

[1] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1627 –1645, sept. 2010, version 3.1.

[2] Navneet Dalal and William Triggs, “Histograms of oriented gradients for human detection,” *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR ’05)*, vol. 1, no. 3, pp. 886–893, 2004.

[3] David G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[4] M. Reisert and H. Burkhardt, “Equivariant holomorphic filters for contour denoising and rapid object detection,” *Image Processing, IEEE Transactions on*, vol. 17, no. 2, pp. 190 –203, Feb. 2008.

[5] Yongxia Zuo, Guoqiang Wang, and Chuncheng Zuo, “Wavelet packet denoising for pavement surface cracks detection,” in *Computational Intelligence and Security, 2008. CIS ’08. International Conference on*, dec. 2008, vol. 2, pp. 481 –484.

[6] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in *Proceedings of the Sixth International Conference on Computer Vision (ICCV 1998)*, New Delhi, India, Jan. 1998, pp. 836–846, IEEE Computer Society.

[7] M. Everingham, A. Zisserman, C. K. I. Williams, and L. Van Gool, “The PASCAL Visual Object Classes Challenge 2006 (VOC2006) Results,” http://www.pascal-network.org/challenges/VOC/voc2006/results.pdf.