SVM-BASED PARAMETER SETTING OF SELF-QUOTIENT ε-FILTER AND ITS APPLICATION TO NOISE ROBUST HUMAN DETECTION

Mitsuharu Matsumoto

The University of Electro-Communications, 1-5-1, Chofugaoka, Chofu-shi, Tokyo, 182-8585, Japan

Keywords: Parameter setting, Nonlinear filter, Support vector machine, Self-quotient ε-filter, Human detection.

Abstract: This paper describes SVM-based parameter setting of self-quotient ε-filter (SQEF), and its application to noise robust human detection combining SQEF, histograms of oriented gradients (HOG), and support vector machine (SVM). Although human detection combining HOG and SVM is a powerful approach, as it uses local intensity gradients, it is difficult to handle noise corrupted images. On the other hand, although human detection combining SQEF, HOG and SVM can realize noise robust human detection, SQEF requires manual parameter setting. Our aim is not only to train SVM but also to adjust the parameter of self-quotient ε-filter using the trained SVM in training procedure. The experimental results show that we can realize noise robust human detection by using SQEF with the obtained parameter, HOG and SVM trained by intact images without noise.

1 INTRODUCTION

Detecting human from images is an important application in image processing. The important requirement is to extract the feature from the images clearly, even in backgrounds under different illumination. Histogram of Oriented Gradients (HOG) algorithm is a useful approach to match this requirement (Dalal and Triggs, 2005). It can extract the feature clearly compared to other existing feature sets including wavelets (Viola et al., 2003). The approach is related to edge orientation (Freeman et al., 1996), SIFT descriptors (Lowe, 2004) and shape contexts (Belongie et al., 2001). Although locally normalized HOG detectors are attractive approaches to detect the human from the image, it is difficult to detect them from the noise corrupted images because it uses local intensity gradients.

To handle the problems, we introduce self-quotient ε-filter (SQEF), which is an advanced noise robust self-quotient filter (SQF) and propose a noise robust SVM-based human detection combining SQEF and HOG.

SQEF (Matsumoto, 2010a; Matsumoto, 2010b) is based on the idea of SQF (Wang et al., 2004) and ε-filter (Arakawa and Okada, 2005).

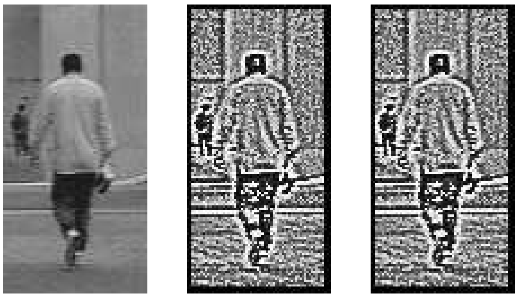
SQF is a simple nonlinear filter to extract the feature from an image (Wang et al., 2004). It needs only an image, and can extract intrinsic lighting invariant property of an image, while removing extrinsic factor corresponding to the lighting. Feature extraction by SQF is simpler than that based on multi-scale smoothing (Gooch et al., 2004). SQF can extract the outline of the objects independent of shadow region. However, as it assumes that the image does not include noise, it can not extract the shape and texture when the noise damages the image. The noise influence becomes large due to the self-quotient effect of SQF.

Although many studies have been reported to reduce the small amplitude noise while preserving the edge (Himayat and Kassam, 1993; Tomasi and Manduchi, 1998), it is considered that ε-filter is a promising approach due to its simple design. It does not need to have the signal and noise models in advance. It is easy to be designed and the calculation cost is small because it requires only switching and linear operation. We can clearly extract the feature from noise corrupted image images by defining SQEF as the ratio of two different ε-filters, and can reduce the noise influence by employing SQEF as preprocessing of HOG.

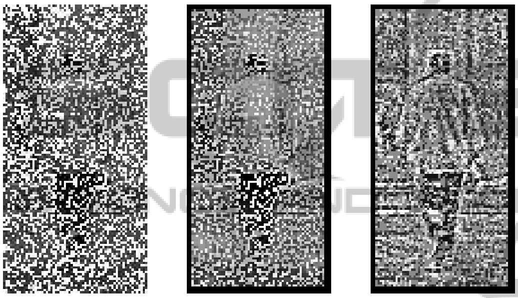
Although human detection combining HOG, SQEF and SVM can realize noise robust human detection, SQEF requires manual parameter setting. Our aim in this paper is not only to train SVM but also to adjust the parameter of ε-filter using the trained SVM in training procedure.

The rests of this paper are organized as follows:

Matsumoto M..
SVM-BASED PARAMETER SETTING OF SELF-QUOTIENT e-FILTER AND ITS APPLICATION TO NOISE ROBUST HUMAN DETECTION.
DOI: 10.5220/0003177102900295
In *Proceedings of the 3rd International Conference on Agents and Artificial Intelligence* (ICAART-2011), pages 290-295
ISBN: 978-989-8425-40-9



(a) A sample image from MIT pedestrian database (file name: per00002.pgm)
(b) Filter output of SQF when we used original image
(c) Filter output of SQEF when we used original image



(d) Impulse noise corrupted image (40% impulse noise)
(e) Filter output of SQF when we used impulse noise corrupted image
(f) Filter output of SQEF when we used impulse noise corrupted image

Figure 1: Self-quotient filter and self-quotient ε-filter from original image and impulse noise corrupted image.

In section 2, we introduce the algorithm of SQEF and describe its features compared to SQF. In section 3, we explain SVM-based parameter setting of self-quotient ε-filter, and implement it in human detection combining SQEF, HOG and SVM. In section 4, experimental results are given to clarify the validness of our approach. Human detection combining SQEF, HOG and SVM with the obtained parameter is also shown with the results using other approaches. A libsvm (Chang and Lin, 2001), MIT pedestrian test set (Papageorgiou and Poggio, 2000) and standard image database (SIDBA) are used as a SVM classifier, positive sample images and negative sample images, respectively throughout the experiments. Conclusion follow in section 5.

2 SELF-QUOTIENT ε- FILTER

We first describe the algorithms of self-quotient filter and self-quotient ε-filter, and explain their features to clarify the handling problem. Let us define $x(i_1,i_2)$ as the image intensity at the point $\mathbf{i} = (i_1,i_2)$ in the image. The aim of self-quotient filter is to separate the intrinsic property and the extrinsic factor, and to remove the extrinsic factor (Wang et al., 2004). To solve the problem, self-quotient filter assumes that a smoothed version of an image has approximately the same illumination as the original one. In self-quotient filter, we first calculate the following equation:

$$z(i_1,i_2) = \frac{x(i_1,i_2)}{F[x(i_1,i_2)]}, \quad (1)$$

where $x(i_1,i_2)$ is the original image and F is the smoothing function.

Due to the process of Eq.1, the texture and edge can be extracted because the original image is divided by the smoothed image. However, self-quotient filter assumes that the image does not include the noise. When we consider the noise corrupted image, the noise is reduced in the smoothed images $F[x(i_1,i_2)]$, while the original image $x(i_1,i_2)$ includes the noise. As a result, the influence from the noise in SQF is emphasized very much due to the self-quotient effect of SQF in Eq.1.

A simple idea to solve the noise influence in SQF is to use two smoothed filters instead of original image as follows:

$$y(i_1,i_2) = \frac{F_1[x(i_1,i_2)]}{F_2[x(i_1,i_2)]}. \quad (2)$$

F_1 and F_2 should be different because the output always becomes 1 if F_1 and F_2 are the same smoothed filter.

However, even if we design SQF by using two different smoothed filters, not only the noise is smoothed but also the texture and shape are blurred. As the blur level of one smoothed filter is different from the other, it is also difficult to handle impulsive noise. Hence, we need to employ alternative filters, which can reduce the small amplitude noise effectively, while preserving the texture and shape information instead of simple smoothed filter. The alternative filters should be simple to keep the simplicity of SQF.

Based on the above prospects, self-quotient ε-filter (SQEF) is designed as follows:

$$y(i_1,i_2) = \frac{\Phi_{\varepsilon_1}[x(i_1,i_2)]}{\Phi_{\varepsilon_2}[x(i_1,i_2)]}, \quad (3)$$

where Φ_ε represents ε-filter described as follows:

$$z(i_1,i_2) = \Phi_\varepsilon[x(i_1,i_2)] = x(i_1,i_2) + \sum_{j_1=-K}^{K}\sum_{j_2=-K}^{K} a(j_1,j_2)F(x(i_1+j_1,i_2+j_2) - x(i_1,i_2)), \quad (4)$$

where $a(j_1, j_2)$ represents the filter coefficient. $a(j_1, j_2)$ is usually constrained as follows:

$$\sum_{j_1=-K}^{K} \sum_{j_2=-K}^{K} a(j_1, j_2) = 1. \quad (5)$$

$F(x)$ is the nonlinear function described as follows:

$$|F(x)| \leq \varepsilon : -\infty \leq x \leq \infty, \quad (6)$$

where ε is a constant number constrained as follows.

$$0 \leq \varepsilon. \quad (7)$$

It should be noted that calculation cost of ε-filter is small because it requires only switching and linear operation. See the references (Arakawa and Okada, 2005) if the reader would like to know the details about ε-filter.

When we apply SQEF to impulse noise corrupted image, it is considered that both ε-filters in SQEF keep the impulse noise in the image unlike when two smoothed filters are employed. Hence, when one filter output in SQEF is divided by the other filter in SQEF, the impulse noise effect is reduced by the self-quotient effects.

Some examples are shown to clarify the difference between self-quotient filter (SQF) and SQEF. Figure 1 shows the examples of filter output of SQEF to show its robust feature extraction from noise corrupted images. We also show the filter output of self-quotient filter (SQF). Fig.1(a) shows a sample image from MIT pedestrian database (Papageorgiou and Poggio, 2000) . Figs.1(b) and 1(c) show the filter outputs of SQF and SQEF, respectively when we used the original image. On the other hand, Fig.1(d) shows the sample image corrupted with 40% impulse noise. Figs.1(e) and 1(f) show the filter outputs of SQF and SQEF, respectively when we used the impulse noise corrupted image. As shown in Fig.1, both SQF and SQEF can extract the feature from the original image. However, SQF cannot extract its feature from the impulse noise corrupted image, while SQEF can extract the feature from the impulse noise corrupted image.

3 SVM-BASED PARAMETER SETTING

This section gives the algorithm of SVM-based parameter setting, and describes the implementation of human detection combining SQEF, Histograms of Oriented Gradients (HOG) and support vector machine (SVM). As is described in the previous section, SQEF can extract the features not only from the intact images without noise but also from the noise corrupted images. However, SQEF requires parameter setting to obtain the adequate filter outputs.

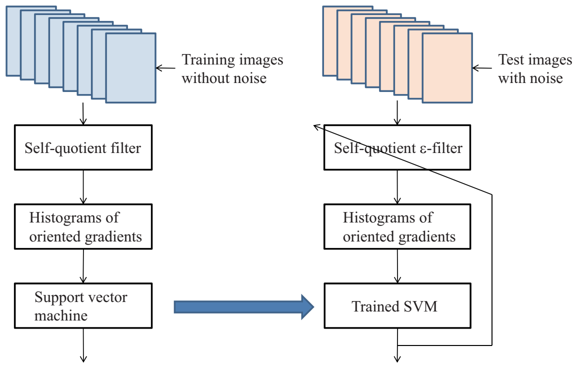


Figure 2: Basic concept of SVM-based parameter setting.

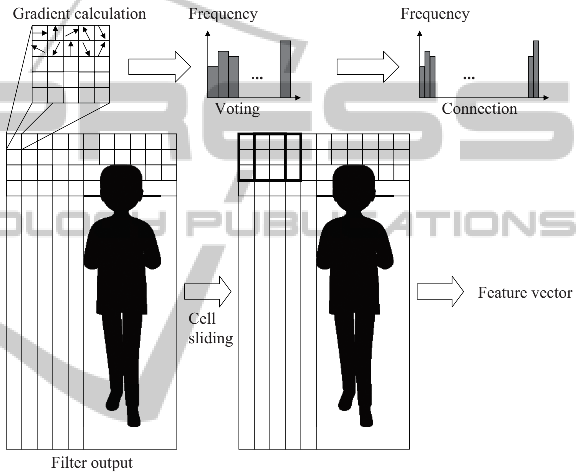


Figure 3: Procedure of Histogram of Oriented Gradients (HOG) from SQEF output.

Figure 2 shows the procedure of our approach to set the parameter of SQEF by using SVM. We first prepare numerous training images without noise, and apply not SQEF but SQF to the training images. We then extract the features from the filter output of SQF by using HOG. Figure 3 shows the procedure of feature extraction from SQF outputs using HOG. The method is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid. When we employ the intact images without noise, as local object appearance and shape are kept in SQF output, the gradient intensity and the gradient direction of SQF are calculated for all the pixels as follows:

$$f_{i_1}(i_1, i_2) = y(i_1+1, i_2) - y(i_1-1, i_2) \quad (8)$$

$$f_{i_2}(i_1, i_2) = y(i_1, i_2+1) - y(i_1, i_2-1) \quad (9)$$

$$m(i_1, i_2) = \sqrt{f_{i_1}^2 + f_{i_2}^2} \quad (10)$$

$$\theta(i_1, i_2) = \arctan \frac{f_{i_2}}{f_{i_1}} \quad (11)$$

Table 1: Experimental results of human detection from impulse noise corrupted image.

Used images	Recognition rate
Original images	50%
Self-quotient filter	50.1%
Self-quotient ε-filter	80.7%

The basic idea of HOG is that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions (Dalal and Triggs, 2005). In practice, this is implemented by dividing the filter output into small spatial regions ("cells"), for each cell accumulating a local 1-D histogram of gradient directions or edge orientations over the pixels of cell. The obtained direction θ ($0^\circ \leq \theta \leq 180^\circ$) is divided with 20° intervals. 9 dimensional feature vector is generated by adding the gradient intensity $m(i_1,i_2)$. We then regard 3×3 cells as "Block" and generate many blocks by sliding on a pixel to pixel basis. The feature vector is finally obtained by combining all the feature vector. The obtained feature vector is used to train support vector machine (SVM). The above procedure corresponds to the left procedure in Fig.2.

After the above procedure, we set the parameter of ε-filter using the trained SVM. The test images with noise are first applied to self-quotient ε-filter. We then extract the feature from the filter outputs of SQEF by using HOG the same as the preprocessing. The obtained feature vector is adopted to SVM trained in the preprocessing. It is expected that the recognition rate will be high if the filter output of SQEF is similar to the filter output of SQF from intact images without noise. Hence, the parameter is obtained as the parameter, which maximizes the recognition rate with regard to the test images. Finally, we can detect human from noise corrupted images by using SQEF with the obtained parameter, HOG and SVM trained by the intact images without noise.

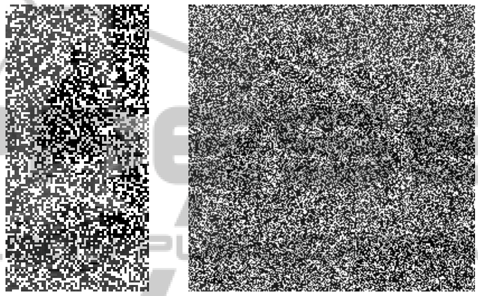
Let us test our criterion experimentally.

4 EXPERIMENTS

We conducted some experiments on SVM-based parameter setting of self-quotient ε-filter. MIT pedestrian database and SIDBA were employed as image database. MIT pedestrian database contains 900 images. The size is 64 pixel $\times$ 128 pixel. Some non person images were selected from standard image database (SIDBA). 900 64 pixel $\times$ 128 pixel images were cut from them. We also prepared 40% impulsive noise corrupted images by adding the impulse noise to the above 1800 images. Figure 4 shows original person / non-person images and its noise-corrupted version. Our aim is to detect human from these types of noise corrupted images not by using the data trained by the impulse noise corrupted image but by using the data trained by intact images without noise.

(a) Person image from MIT pedestrian database (per00003.pgm)

(b) Non-person image from SIDBA (Airplane)



(c) Person image from MIT pedestrian database with 40% impulse image (per00003.pgm)

(d) Non-person image from SIDBA with 40% impulse noise (Airplane)

Figure 4: Sample images of person image and non-person image (Original and noise corrupted images).

As a SVM tool, we used libsvm, a library for support vector machines (Chang and Lin, 2001), and employed default setting and parameters throughout the experiments for simplicity. In the experiments, we used original 450 pedestrian images from MIT pedestrian database and 450 non-person images from SIDBA. SVM was trained by using the data combining SQF and HOG from the above 900 person / non-person images. To simplify the experiments, we set ε_1 was set to 0, (Original image), and tried to set ε_2 by using the proposed method.

We then checked the relation between ε value and recognition rate when we used the method combining SQEF with changing the parameter, HOG and SVM trained by the previous procedure. Figure 5 shows the

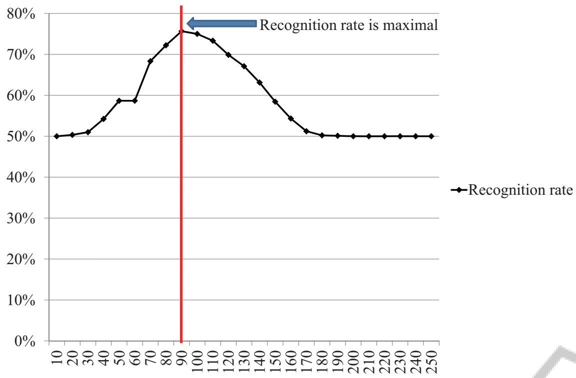


Figure 5: Relation between ε value and recognition rate: Impulse noise corrupted images were used.

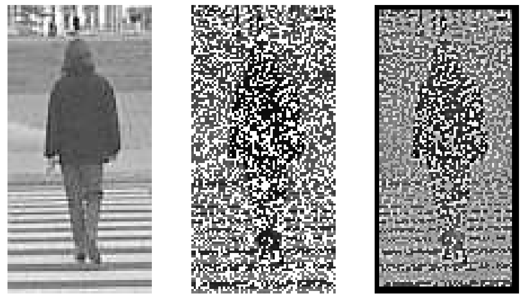
relation between ε value and recognition rate. The parameter, which maximizes the recognition rate was 90 as shown in Fig. 5.

Finally, we conducted the experiments of human detection by using SQEF with the obtained parameter, HOG and SVM. The test images are the remaining 450 pedestrian images from MIT pedestrian database and the remaining 450 non-person images from SIDBA with impulse noise, which are different from the training images. For comparison, we also tested to classify them using the method combining HOG and SVM, and the method combining SQF, HOG and SVM. Table 1 shows the recognition results. It should be note that the recognition rate becomes 50% even if the system always says "human" or "non-human" because the sample images include human and non-human images evenly. In other words, the recognition results of the comparison method were almost no meaning when the noise were added. On the other hand, the proposed approach could detect human from noise corrupted images over 80% using training data with intact images without noise and the obtained parameter.

We finally show the example of the filter output of self-quotient ε-filter with the obtained parameter. Figure 6 shows the obtained results. Figure 6(a) shows a sample image from MIT pedestrian database. Figure 6(b) shows the sample image corrupted with impulse noise. Figure 6(c) shows the filter output of SQF when we used the sample image corrupted with impulse noise.

Figure 6(d) shows the filter output of SQEF with the obtained parameter (ε_2=90) when SQEF is applied to the sample image corrupted with impulse noise. For comparison, we also show the filter outputs of SQEF with regard to the sample image corrupted with impulse noise when ε_2 was set to 10 and 250 as shown in Figs 6(e) and Fig. 6(f), respectively.

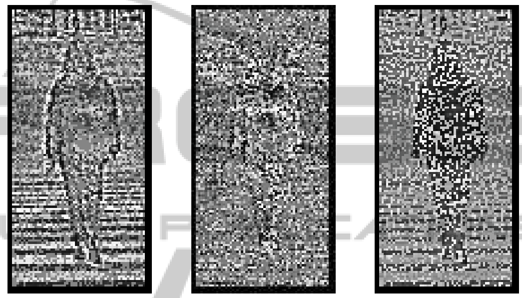
As shown in Fig.6, SQEF could extract the feature



(a) A test image from MIT pedestrian database (file name: per00927.pgm)

(b) Impulse noise corrupted image (40% impulse noise)

(c) Filter output of SQF when we used impulse noise corrupted image



(d) Filter output of SQEF when we used impulse noise corrupted image. (ε_2=90)

(e) Filter output of SQEF when we used impulse noise corrupted image. (ε_2=10)

(f) Filter output of SQEF when we used impulse noise corrupted image. (ε_2=250)

Figure 6: Self-quotient image and self-quotient ε-filter from original image and impulse noise corrupted image.

from the impulse noise corrupted image with the obtained parameter, while it could not extract the feature from the impulse noise corrupted image with inadequate parameters.

5 CONCLUSIONS

In this paper, we proposed SVM-based parameter setting of self-quotient ε-filter (SQEF) and implement it to human detection combining SQEF, HOG and SVM. We conducted some experiments and compared the filter output with the adequate parameter to the filter outputs with other parameters. Throughout the experiments, the proposed method could obtain the adequate parameter and could realize noise robust human detection from noise corrupted images using the training data with the clean image without noise. For future works, we would like to apply our approach to robot vision.

ACKNOWLEDGEMENTS

This research was supported by the research grant of Support Center for Advanced Telecommunications Technology Research (SCAT), by the research grant of Foundation for the Fusion of Science and Technology, by Special Coordination Funds for Promoting Science and Technology, and by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), 20700168, 2008.

REFERENCES

Arakawa, K. and Okada, T. (2005). ε-separating nonlinear filter bank and its application to face image beautification. In *IEICE Transactions on Fundamentals*, pages 1216–1225. IEICE.

Belongie, S., Malik, J., and Puzicha, J. (2001). Matching shapes. In *Proc. of Int'l Conf. on Computer Vision*, pages 454–461. IEEE.

Chang, C.-C. and Lin, C.-J. (2001). LIBSVM: a library for support vector machines. In *http://www.csie.ntu.edu.tw/ cjlin/libsvm*.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proc. of Int'l Conf. on Computer Vision and Pattern Recognition.*, pages 886–893. IEEE.

Freeman, W. T., Tanaka, K., Ohta, J., and Kyuma, K. (1996). Computer vision for computer games. In *Proc. of Int'l Conf. on Automatic Face and Gesture Recognition*, pages 100–105. IEEE.

Gooch, B., Reinhard, E., and Gooch, A. (2004). Human facial illustrations: Creations and psychological evaluation. In *ACM transactions on Graphics*, pages 27–44. ACM.

Himayat, N. and Kassam, S. (1993). Approximate performance analysis of edge preserving filters. In *IEEE Trans. on Signal Processing.*, pages 2764–2777. IEEE.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. In *Int'l Journal of Computer Vision*, volume 60, pages 91–110. Springer.

Matsumoto, M. (2010a). Feature extraction from noisy face image using self-quotient ε-filter. *Proc. of Int'l Conf. on Computer Engineering and Technology*, pages 395–399.

Matsumoto, M. (2010b). Self-quotient ε-filter for feature extraction from noise corrupted image. *IEICE Trans. on Infomation and Systems*.

Papageorgiou, C. and Poggio, T. (2000). A trainable system for object detection. In *Int'l Journal of Computer Vision*, volume 38, pages 15–33. Springer.

Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Proc. of Int'l Conf. on Computer Vision.* IEEE.

Viola, P., Jones, M. J., and Snow, D. (2003). Detecting pedestrians using patterns of motion and appearance. In *Proc. of Int'l Conf. on Computer Vision*, volume 1, pages 734–741. IEEE.

Wang, H., Li, S. Z., and Wang, Y. (2004). Face recognition under varying lighting conditions using self quotient image. In *Proc. of Int'l Conf. on Automation Face and Gesture Recognition*. IEEE.