Optimization of Weighted Vector Directional Filters Using an Interactive Evolutionary Algorithm

Joseph James Mist
School of Physical Sciences, University of Kent,
Canterbury, Kent, United Kingdom, CT2 7NH
jm441@kent.ac.uk

Stuart James Gibson
School of Physical Sciences, University of Kent,
Canterbury, Kent, United Kingdom, CT2 7NH
s.j.gibson@kent.ac.uk

ABSTRACT

Weighted vector directional filters are used to enhance multichannel image data and have attracted a lot of interest from researchers in the image processing community. This paper describes a novel method for deriving the weights of a vector directional filter that uses an interactive evolution strategy. We performed an empirical study in which 30 participants each developed two filters using our approach. Each participant compared the performance of his/her filters to the basic vector directional filter and a filter that had previously been developed using a genetic algorithm. Of the filters studied, our interactive approach was the most effective at removing salt and pepper noise for the case when the percentage of corrupt image pixels was low.

Categories and Subject Descriptors

G.1.6 [**Numerical Analysis**]: Optimization—*global optimization*; I.4.3 [**Image Processing and Computer Vision**]: filtering

Keywords

vector directional filter; interactive evolutionary algorithm; evolution strategy; image quality measure; spatial filters

1. INTRODUCTION

Interpretation of image data by human analysts is prevalent in many fields including medical imaging, astronomical imaging, and the biological sciences. An important, and frequently adopted, preliminary step in the interpretation process is to enhance the image of interest using standard image processing techniques thereby assisting in the objective. The optimal image processing technique and associated parameter settings can depend on a number of factors such as the extent and type of any degradation, the particular image, or the purpose of the image. The specific nature of these factors may not be known in advance. In such circumstances it is reasonable to assume that a human analyst should be able to adapt the image enhancement process to suit a specific goal.

The aims of our work are: 1) To optimize the weights of weighted vector directional filters (WVDFs) using an interactive evolutionary algorithm (IEA). 2) To test the robustness of this approach to variations in noise level and compare it with other WVDFs. 3) To compare image quality measures (IQMs) with human perceptions of image quality in relation to the task of removing salt and pepper noise from images.

The remainder of the paper is organized into four sections. Section 2 covers the relevant theory of IEAs, vector directional filters, and IQMs. Section 3 describes our experimental method. Our results and interpretation of our experimental data are included in Section 4. We discuss the merits of our approach in Section 5.

2. THEORY

2.1 Interactive evolutionary algorithms

In an evolutionary algorithm (EA) a fitness function is used to measure the suitability of candidate solutions to a problem. In an IEA the fitness of the candidate solutions is determined by subjective choices made by the human user. Using human assessment to perform the role of a fitness function places limitations on IEAs that are not an issue for EAs. Typically, EAs can evaluate the fitness of thousands of individuals per generation over hundreds of generations. Conversely, in IEAs the population size and number of generations are crucial factors in the design of the algorithm to avoid user fatigue and to accommodate time constraints.

In our work we reduce user fatigue by choosing an evolutionary algorithm that requires the user to select a single member of the population to seed the following generation. This feature of the design dictates the use of evolution strategy (ES) because genetic algorithms (GAs) apply a crossover operator to the genes of more than one parent. The main evolutionary operators used in an ES are mutation and selection.

2.2 Vector filters

A common image processing task is to remove noise from an image. One approach is to use a WVDF. The optimal weights for the WVDF are difficult to determine analytically. To address this problem Lukac et al. [7] applied a GA to adapt the weights to match varying image and noise characteristics. In Lukac et al.'s work the GA develops a filter on a training image which is then used to denoise previously

GECCO'13 Companion, July 6–10, 2013, Amsterdam, The Netherlands.

unseen images. The training image was deliberately contaminated with noise of a known profile and was assumed to be previously noise-free. The efficacy of a filter can then be measured by comparing the filtered image to the original uncontaminated image.

In this paper we study images contaminated with salt and pepper noise. Salt and pepper noise is normally treated using order statistic filters. The application of (scalar) order statistic filters in gray-scale image processing has been studied extensively. However, the extension of the scalar theory to multichannel data is problematic because the process requires an ordering in multidimensional space, which is not well defined. One approach is to perform the ordering on each of the multispectral channels independently thereby treating each of the channels as a separate gray-scale image. A more elegant approach that has been developed for noise reduction uses vector order statistics. The best known of the vector order filters is the vector median filter [1] with more sophisticated filters including the basic vector directional filter (BVDF) and the WVDF. A general overview of vector filtering can be found in [4]. We chose to work with WVDFs because Lukac et al.'s work on using a GA to optimize filter weights also used WVDFs [7] and so provided us with an opportunity to compare the performance of an EA to that of an IEA.

In the vector filtering approach pixel values are represented as points in a three-dimensional space. Vectors are then constructed from the origin to these points. The axes of the three-dimensional space correspond to the red, green, and blue color channels. In vector directional filtering, the distance between two colors is defined as the angle between their corresponding vectors as measured from the origin of the red, green, blue color space. The output of the filtering window is the pixel that has the smallest weighted sum of the angles between itself and the other pixels in the window. The filtering window W is n pixels in size (in our case W is a 3×3 window so $n = 9$). The weight of the i-th position in the window is w_i. Using a vector representation, the i-th and j-th pixels in the window are $\mathbf{x}_i$ and $\mathbf{x}_j$ respectively. The output $\mathbf{x}_{\mathrm{WVDF}}$ is the vector that satisfies

$$\mathbf{x}_{\mathrm{WVDF}} = \underset{\mathbf{x}_j \in W}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \arccos\left(\frac{\mathbf{x}_i \cdot \mathbf{x}_j}{\|\mathbf{x}_i\| \, \|\mathbf{x}_j\|}\right). \tag{1}$$

2.3 Image quality measures

IQMs are used to assess the effectiveness of various image compression and restoration techniques. The most commonly used IQMs are full reference IQMs. Full reference IQMs compare a clean original image to one that has undergone compression or has been corrupted and then restored. The more similar the processed image is to the original, the more effective the process is judged to be.

Non-interactive EAs are generally quicker and more convenient to use than IEAs but require a fitness function which, in the context of perceptual image enhancement, cannot be defined mathematically. Therefore an appealing idea is to emulate perceptual image quality using an IQM. Whilst the development of IQMs and the assessment of their relative performances has been studied [2], little work has been done to compare them with human evaluations of image quality. In this paper we compare five IQMs with human evaluation of image quality when considering the task of denoising.

The IQMs that we compared to human evaluations of image quality were the mean absolute error (MAE), mean square error (MSE), mean quartic error (MQE), normalized color index (NCD)[8], and structural similarity index (SSIM)[10].

3. METHOD

3.1 Our IEA

In our work each filter is represented as a chromosome consisting of ten genes. Nine of the genes are the filter weights $w_1, \ldots, w_9$, are real coded, and have values in the range $0 \leq w_i \leq 1$. The tenth gene is the mutation step size, σ, and is subject to the condition $\sigma \geq 0.075$.

We used uncorrelated mutation with one step size for the mutation component of the algorithm. During the mutation stage, the step size gene is mutated by

$$\sigma' = \sigma \cdot e^{\frac{N(0,1)}{2}}$$

where σ' and σ are the new and old step sizes respectively, $N(0,1)$ is a number taken at random from the standard normal distribution. The step size parameter is self adapting, the step size is evolved alongside the filter weights. The theory is that appropriate step sizes are more likely to generate desirable filter weights. The filter weights are then mutated by

$$w_i' = w_i + \sigma' \cdot N(0,1)$$

where w_i' and w_i are the new and old weights respectively at position i in the window. The mutation component of the algorithm is described in more detail in [5].

3.2 Test images

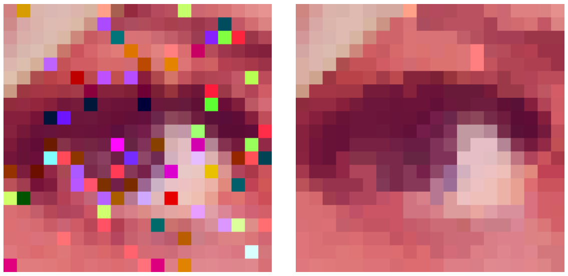
We used the popular Lena test image scaled to a size of 256×256 pixels. Two contaminated versions of the Lena image, which we refer to as I1 and I2, were created. Salt and pepper noise was applied to each of the red, green, and blue channels of each image. The probability of a particular pixel being contaminated on any channel was 2% for I1 and 8% for I2.

3.3 Optimization of filter weights

For this experiment we used 30 participants. The number of participants and treatments considered in our study is similar to previous work in the field of interactive evolutionary computation [9, 3]. Repetitive tasks quickly lead to fatigue and this was the central consideration in our experimental design. Accordingly, each participant was given the task of developing two filters to remove noise from images using our IES. The participants were asked to continue developing their filters until they were satisfied with the result or they thought no further improvement was possible. The participants were not shown the original noise-free image at any point during the experiment.

A compromise is sought between the number of images shown to the participant per generation and the effort required to compare them. We have learned from experience derived from real world applications based on our previous work [6] that the human visual system becomes overwhelmed if required to compare too many images at once. We found four images per generation to be appropriate for the task of denoising.

Figure 1: A section of the 8% noise image (a) and after applying a participant generated filter (b)



(a) Before (b) After

The participants were required to give reasons for their choices of images; for example "Images 1 and 3 have less noise than 2 and 4, image 1 has an annoying pixel on Lena's nose hence I choose image 3." This information was useful for explaining the performances of the IQMs.

The initial population in each run consisted of four mutants of the identity filter ($w_5 = 1, w_i = 0$ for other values of i). We chose the identity filter as the starting point because of its mathematical simplicity and to encourage the development of filters that were effective at removing noise yet introduced few filter artifacts. The initial step sizes for the first generation were drawn at random from $U(0, 1)$. Every generation thereafter consisted of the fittest (selected) filter of the previous generation and three mutant offspring spawned from it.

3.4 Assessment of IEA filter performance

After developing their filters, participants were asked to compare the performance of four filters: their own filter developed on I1 (which we call F1), their own filter developed on I2 (which we call F2), the BVDF, and the GA optimized W2 filter (which has weights 0.1526, 0.2610, 0.2007, 0.2059, 1, 0.1992, 0.2115, 0.2581, 0.1435) developed by Lukac et al. [7]. The four filters were applied to each of the images in turn and the results displayed in a 2×2 image array similar to the one used for developing the filters. The participant was asked to rank the images in order of image quality, giving reasons for their preferences as they did when developing their filters.

4. RESULTS

4.1 Analysis of ranked performance

The relative performance of the four filters was assessed using pair wise comparisons based on the ranked data obtained by the method described in Section 3.4. The number of times each filter was preferred to another filter was counted over all of the participant rankings. The result is shown in Table 1.

To determine the statistical significance of our results we applied a two-tailed binomial test to each of the pair wise preference counts. The null hypothesis is that for each entry in Table 1 there is no significant deviation from 15 counts.

For I1 the pair wise comparisons show that W2 performed significantly worse ($p < 0.001$) on image I1 than every other filter. None of the other pairwise comparisons are statistically significant. The total number of counts (all participants) recorded for each filter indicated the following the order of preference (best to worst): F1, F2, BVDF and W2. Hence the consensus was that participants preferred their own IEA filter (optimized for I1) to all other filters.

For I2 the order of preference was BVDF, F2, F1 and W2 with each pair wise comparison resulting in a statistically significant difference ($p < 0.005$). Therefore, for 8% noise the IEA developed filters were not as successful as the BVDF. In fact the BVDF was judged to be significantly better ($p < 0.001$) than all other filters and again W2 performed poorly in each pair wise comparison ($p < 0.001$). An example of a user generated filter is given in Figure 1.

Since F1 was optimized for I1 we expected it to perform better than F2 on I1. Conversely, F2 should be more effective than F1 on I2. This assertion was tested by collating the results for F1 and F2 from Table 1(a) and Table 1(b) and applying a one-tailed binomial test. The total number of counts in favor of the image specific IEA filter was 41 out of a possible 60 which was significant at $p < 0.005$. Therefore we conclude that our IEA was able to adapt to the differing levels of noise in the images.

Table 1: Filter preferences

(a) I1: 2% noise

	Preferred filter			
Less favored filter	F1	F2	BVDF	W2
F1	—	11	12	5
F2	19	—	14	5
BVDF	18	16	—	5
W2	25	25	25	—
Total	62	52	51	15

(b) I2: 8% noise

	Preferred filter			
Less favored filter	F1	F2	BVDF	W2
F1	—	22	29	1
F2	8	—	27	1
BVDF	1	3	—	1
W2	29	29	29	—
Total	38	54	85	3

4.2 Objective measures of image quality

To assess the efficacy of the IQMs we calculated the Spearman rank correlation coefficient (Spearman's ρ) between the ranks each participant assigned to the filtered images and the ranks assigned based on the IQMs for the same filtered images. We calculated the means of the ρ values over all of the participants (see Table 2). Given that $\rho = 0.8$ indicates that an adjacent pair filters exchange places in the rank order, we considered an IQM with a mean ρ of 0.8 to be a reasonable model of human evaluation.

From Table 2 it can be seen that the MAE is a poor IQM for modeling human evaluation of filter performance. For I1 none of the IQMs provided a satisfactory model of human evaluation, although the MQE was much better than the other IQMs. As the MQE significantly outperforms the MSE

Table 3: Mentions of image selection considerations when developing and ranking the filters

Consideration	I1: 2% noise Developing	I1: 2% noise Ranking	I2: 8% noise Developing	I2: 8% noise Ranking
General noise	30	29	30	30
Single pixel	11	9	9	3
Filter artifacts	16	20	17	10

Table 2: Mean Spearman ρ coefficients

Image	MAE	MSE	MQE	NCD	SSIM
I1: 2% Noise	-0.233	-0.287	0.367	-0.267	-0.253
I2: 8% Noise	-0.387	0.893	0.880	0.687	0.767

on I1 and was nearly as good as the MSE on I2 we concluded that of the five IQMs we tested the MQE provides the best model of human evaluation when measuring the efficacy of the filters for removing salt and pepper noise.

4.3 Selection considerations

When the participants were explaining the reasons for their image selections, nearly all of their comments could be divided into three categories: general noise, single pixels, and filter artifacts. General noise refers to many noisy pixels, either over the entire image or a particular part (e.g. Lena's face). Single pixel refers to a particular noisy pixel that the participant has noticed, generally a pixel in an otherwise noise free part of the image and often in a prominent place such as on Lena's nose. Filter artifacts refer to parts of the image that have been worsened because uncontaminated pixels have been altered by the filter. Table 3 shows a count of the number of participants who mentioned each of the three considerations when developing their filters and when ranking the filters.

It can be seen from Table 3 that general noise was the most important consideration for the participants. Single pixels were a more important consideration to the participants when ranking the filters applied to I1 than when ranking the filters applied to I2. We believe that this was because noisy regions in I1 were more likely to to contain only a single noisy pixel after filtering than was the case for I2. The difference between the number of participants who cited filter artifacts as a consideration at the ranking stage can be explained by the fact that as I2 was a noisier image, more of the participants found the introduction of filter artifacts to be of less concern than the removal of noise.

5. CONCLUSION

We have demonstrated that the weights of a vector directional filter can be obtained using a simple IEA in which assessments of image quality are made by a human user. Our method was more effective for improving perceptual image quality than a filter previously developed using a GA. In the presence of 2% salt and pepper noise, the IEA filter was also more successful at improving image quality than the well known BVDF.

The poor performance of the GA based filter in our study can be attributed to the use of the MAE for the optimization of its fitness function. We evaluated five objective IQMs and found that whilst the MSE and MQE provided a good model of human opinion on the noisier image, none of the image quality measures were satisfactory for both 2% and 8% noise. Of the five IQMs, the MAE was least similar to human perception of image quality. The poor performance of the IQMs provides evidence to support the use of human evaluation and the IEA approach.

We intend to extend our work to a wider range noise models thereby making the technique more suitable for use in camera equipped consumer technology.

6. REFERENCES

[1] J. Astola, P. Haavisto, and Y. Neuov. Vector median filters. *Proc. IEEE*, 78(4):678–689, Apr. 1990.

[2] I. Avcıbaş, B. Sankur, and K. Sayood. Statistical evaluation of image quality measures. *J. Electron. Imag.*, 11(2):206, Apr. 2002.

[3] R. Breukelaer, M. Emmerich, and T. Bäck. On interactive evolution strategies. In F. R. et al, editor, *Appl. Evol. Comput., EvoWorkshop2006: EvoINTERACT*, volume 3907, pages 530–541. Springer-Verlag GmbH LNCS, 2006.

[4] M. Cree. Observations on adaptive vector filters for noise reduction in color images. *IEEE Signal Process. Lett.*, 11(2):140–143, Feb. 2004.

[5] A. E. Eiben and J. E. Smith. *Introduction to Evolutionary Computing*. Springer-Verlag, Berlin, Germany, 2003.

[6] B. George, S. J. Gibson, M. I. Maylin, and C. J. Solomon. EFIT-V : interactive evolutionary strategy for the construction of photo-realistic facial composites. In *GECCO '08: Proc. 10th Annu. Conf. Genet. Evol. Comput.*, pages 1485–1490, New York, NY, USA, 2008. ACM.

[7] R. Lukac, K. N. Plataniotis, and A. N. Venetsanopoulos. Color image denoising using evolutionary computation. *Int. J. Imag. Syst. Technol.*, 15(5):236–251, Oct. 2005.

[8] K. N. Plataniotis, D. Androutsos, and A. N. Venetsanopoulos. Adaptive fuzzy systems for multichannel signal processing. *Proc. IEEE*, 87(9):1601–1622, Sep. 1999.

[9] J. C. Quiroz, S. J. Louis, and S. M. Dascalu. Interactive evolution of XUL user interfaces. In *GECCO '07: Proc. 9th Annu. Conf. Genet. Evol. Comput.* ACM Press, 2007.

[10] Z. Wang and A. C. Bovik. *Modern Image Quality Assessment*. Morgan and Claypool, San Rafael, 2006.