ANISOTROPIC DIFFUSION BY QUADRATIC REGULARIZATION

Marcus Hund and Bärbel Mertsching
University of Paderborn, Faculty of Computer Science
Electrical Engineering and Mathematics, GET Lab, Germany

Keywords: Edge Preserving Smoothing, Noise Removal, Anisotropic Diffusion, Regularization.

Abstract: Based on a regularization formulation of the problem, we present a novel approach to anisotropic diffusion that brings up a clear and easy-to-implement theory containing a problem formulation with existence and uniqueness of the solution. Unlike many iterative applications, we present a clear condition for the step size ensuring the convergence of the algorithm. The capability of our approach is demonstrated on a variety of well known test images.

1 INTRODUCTION

The idea of anisotropic diffusion in image processing, i. e. diffusion along a preferred orientation, was first introduced by Perona and Malik (Perona and Malik, 1990). Their method was designed to realize image smoothing with simultaneous edge enhancement. In this context an important issue is the relation of the method to the scale space theory as introduced by Witkin. In the last decade, the most important contribution to the topic is the work of Weickert ((Weickert, 1996), (Weickert, 1999)). Like Perona and Malik, Weickert uses the diffusion equation

$$\frac{\partial u}{\partial t} = \nabla \cdot (D \nabla u) \tag{1}$$

and provides a well-founded mathematical framework (Weickert, 1996). He derives the diffusion tensor D from a structure tensor description of the input image in order to describe the local image structure. A major drawback of the structure tensor description, especially combined with a Gaussian regularization, is the high orientation uncertainty in the presence of junctions in the image.
Even though there is a close relationship between diffusion filters in the sense of eq. (1) and regularization methods, one main difference is the fact that the scale space in eq. (1) is given by the time step t, i. e. the process converges to an one valued image representing the maximum smoothed version of the input image. Contrary to this the scale space in regularization-based approaches is given by a smoothing weight within an energy or cost functional that has to be minimized. Although there exist a wide range of regularization-based approaches to image processing (see e. g. (Ito and Kunisch, 1999), (Nordstrom, 1989), (Aubert et al., 1997) and references therein), we are not aware of a regularization-based approach to anisotropic diffusion that brings up a clear and easy-to-implement theory containing a problem formulation with existence and uniqueness of the solution and an iterative scheme with a scale dependent step size that guarantees the convergence of the method. Exactly this is what we present in the following sections. In (Aubert et al., 1997) the superiority of so called edge preserving models over quadratic regularization is stated. Contrary to this, we will show that edge preserving diffusion indeed can be formulated as quadratic regularization.
The most suggesting application of edge preserving regularization is image denoising. In (Portilla et al., 2003) a good overview to existing denoising methods is given. In an extensive quantitative evaluation Portilla et al. show that their approach appears to be the most accurate one compared to other state-of-the-art methods. Therefore we will give a short comparison to their method.

2 GLOBAL OPTIMIZATION

2.1 Problem Formulation

In our approach the discrete image is regarded as a vector $\xi_0 \in \mathbb{R}^{(\mathbf{m}\cdot\mathbf{d})}$ with m being the total number of

Hund M. and Mertsching B. (2008).
ANISOTROPIC DIFFUSION BY QUADRATIC REGULARIZATION.
In *Proceedings of the Third International Conference on Computer Vision Theory and Applications*, pages 101-107

pixels in the image and d the number of colors used. Two assumptions are made about the cost function. First, it should cause costs if the image elements $\xi_{(x,k)}$ differ from the initial pixel values $\xi_{(x,k)_0}$. Second, neighboring elements must satisfy a continuity constraint. This leads to a cost function of the form

$$\begin{aligned} P(\xi) &= \frac{1}{2}\cdot\sum_{k=1}^{d}\sum_{i=1}^{m}(\xi_{(i,k)}-\xi_{(i,k)_0})^2 \\ &\quad +\frac{c}{4}\cdot\sum_{k=1}^{d}\sum_{i=1}^{m}\sum_{j\in U_i} w_{ij}(\xi_{(i,k)}-\xi_{(j,k)})^2 \quad (2) \end{aligned}$$

with $\xi=(\xi_{(1,1)},\ldots,\xi_{(m,1)},\xi_{(1,2)},\ldots,\xi_{(m,d)})^T$ with a smoothness or scaling factor c. Here, U_x is the neighborhood of a given pixel x with a weighting factor $w_{xi}>0$ with $\sum_{i\in U_x} w_{xi}\leq Q:=\sum_{i\in U_x}1$. Note that $w_{xi}\neq w_{ix}$ is allowed. Due to the cost function there must exist a minimum. The minimum ξ must satisfy $\nabla P(\xi)=0$, which leads to a system of linear equations

$$\nabla P(\xi)=\mathbf{A}\cdot\xi-\xi_0=0 \quad (3)$$

with the elements of $\nabla P(\xi)$ given by

$$\begin{aligned} \frac{\partial}{\partial\xi_{(x,k)}}P(\xi) &= \xi_{(x,k)}-\xi_{(x,k)_0} \\ &\quad +\frac{c}{2}\Big(\sum_{j\in U_x} w_{xj}(\xi_{(x,k)}-\xi_{(j,k)}) \\ &\quad +\sum_{\{i|x\in U_i\}} w_{ix}\,(\xi_{(x,k)}-\xi_{(i,k)})\Big) \quad (4) \end{aligned}$$

Consequently, the matrix $\mathbf{A}=(a_{i,j})_{i,j}\in\mathbb{R}^{(m\cdot d)\times(m\cdot d)}$ in (3) is given by

$$a_{i,j}=\begin{cases} 1+\frac{c}{2}(\sum_k w_{ik}+\sum_k w_{ki}) & \text{for} \quad i=j \\ \frac{c}{2}(w_{ji}+w_{ij}) & \text{else} \end{cases} \quad (5)$$

with $w_{kl}=0$ for $l\notin U_k$. Obviously, $\mathbf{A}$ is symmetric and positive definite. Therefore, an inverse $\mathbf{A^{-1}}$ must exist and we have a unique solution of eq. (3).

2.2 Numerical Aspects

Due to round-off errors direct methods are less efficient for large systems than iterative ones. Furthermore, the computation of an inverse would destroy the zeros in the system matrix $\mathbf{A}$ and it would thus lead to a high storage and computation effort. Unfortunately, in many iterative applications no theoretical stability bound for convergence to a unique minimum is available. This results in reducing the step size in an experimental way, until the process remains stable ((Perona and Malik, 1990), (Scharr and Weickert, 2000), (Aubert et al., 1997), (Nordstrom, 1989)). Hence changing the system parameters in these cases may cause serious problems. For the solution of (3), we use the gradient descent method and from this we will deduce a theoretical condition for the stability bound.Starting with $\xi^0\in\mathbb{R}^{(m\cdot d)}$ the vector ξ^k is iteratively updated via

$$\xi^{k+1}=\xi^k-\lambda_k\nabla P(\xi^k), \qquad \lambda_k>0 \quad (6)$$

Instead of using the local optimal step size $\lambda_k=\frac{(\mathbf{r_k},\mathbf{r_k})}{(\mathbf{Ar_k},\mathbf{r_k})}$ with the residuum $\mathbf{r_k}$, which has to be computed for each iteration and therefore causes high computational costs, we use the following considerations for the choice of a constant step size λ. With a given linear system of equations of the form $\mathbf{Ax}=\mathbf{b}$, an iterative scheme $\phi:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}^n$ is called linear, if matrices $\mathbf{M},\mathbf{N}\in\mathbb{R}^{\mathbf{n}}$ exist, such that $\phi(\mathbf{x},\mathbf{b})=\mathbf{Mx}+\mathbf{Nb}$. The process is said to be consistent with the matrix $\mathbf{A}$, if the solution $\mathbf{A^{-1}}$ is a fixed point of ϕ. A linear iteration scheme is consistent with the matrix $\mathbf{A}$ if and only if $\mathbf{M}=\mathbf{I}-\mathbf{NA}$. Furthermore, the scheme is convergent if and only if the spectral radius $\rho(\mathbf{M})$ of the iteration matrix $\mathbf{M}$ satisfies the condition $\rho(\mathbf{M})<1$. If ϕ is convergent and consistent with the matrix $\mathbf{A}$, the limiting value of the sequence $\mathbf{x_k}=\phi(\mathbf{x_{k-1}},\mathbf{b})$ solves $\mathbf{Ax}=\mathbf{b}$ for any choice of $\mathbf{x_0}$.Considering (3) and the iteration scheme (6) with a constant step size $\lambda_k=\lambda$, we receive the linearity and consistence with the matrix $\mathbf{A}$ with $\mathbf{N}=\lambda\mathbf{I}$ and $\mathbf{M}=\mathbf{I}-\mathbf{NA}$. It remains to show under which condition $\rho(\mathbf{M})<1$ is fulfilled. Let $Q_i=\frac{1}{2}\sum_k(w_{ik}+w_{ki})$, again with $w_{kl}=0$ for $l\notin U_k$ and let Q be the maximum number of pixels in a neighborhood U, i. e. $Q>Q_i$. Using $\rho(\mathbf{M})\leq\|\mathbf{M}\|$ for any matrix norm $\|\,.\,\|$ we consider the row-sum norm $\|\,.\,\|_\infty$ and claim $\sum_{k=1}^{m\cdot d}|a_{ik}|<1$ for any row i. This leads to

$$|1-\lambda(1+c\cdot Q_i)|+\lambda\cdot c\cdot Q_i<1 \quad (7)$$

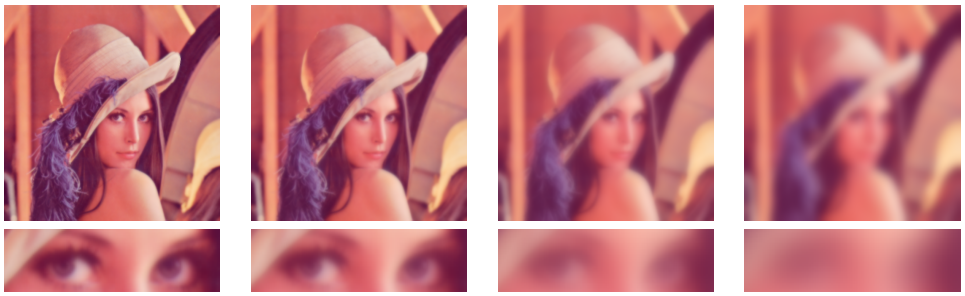
and hence

$$0<\lambda<\frac{2}{1+2\cdot c\cdot Q_i} \quad (8)$$

In the following, we use

$$\lambda:=\frac{1}{1+c\cdot Q}<\frac{2}{1+2\cdot c\cdot Q}\leq\frac{2}{1+2\cdot c\cdot Q_i} \quad (9)$$

We receive a step size that is dependent on the scale, i. e. the smoothness factor c and the maximum number of pixels Q in the neighborhood U. It has to be noted that there is a difference between ξ^0 and ξ_0. The solution is independent from the starting vector ξ^0, but it depends on the initial image ξ_0.



(a) (b) (c) (d)

Figure 1: Linear isotropic scale space for the Lena image: (a) c=1, (b) c=3, (c) c=20 and (d) c=250.

(a) (b) (c) (d)

Figure 2: Anisotropic scale space for the Lena Image with 20% uncorrelated noise for each pixel in the input image: (a) Input image, (b) c=3, (c) c=20 and (d) c=250.

3 DIFFUSION

3.1 Isotropic Diffusion

In the simpliest case we have a quadratic neighborhood U_i of a pixel i and symmetric weighting factors $w_{ij} = w_{ji} = 1$. Eq. (4) then is simplified to

$$\frac{\partial}{\partial \xi_{(x,k)}} P(\xi) = \xi_{(x,k)} - \xi_{(x,k)_0} + c \sum_{j \in U_x} (\xi_{(x,k)} - \xi_{(j,k)}) \quad (10)$$

This is the classic case of isotropic linear diffusion.

The isotropic diffusion as well as anisotropic diffusion with symmetric weights, i. e. $w_{ij} = w_{ji}$, was already used by us in the context of stereoscopic depth estimation ((Hund, 2002), (Brockers et al., 2005)). There, eq. (10) and eq. (6), respectively, were used to eliminate the ambiguities in the stereoscopic disparity space.

Fig. 1 demonstrates the linear isotropic diffusion properties on the Lena image. With growing regularization parameter c the image is isotropically smoothed until all image detail is lost.

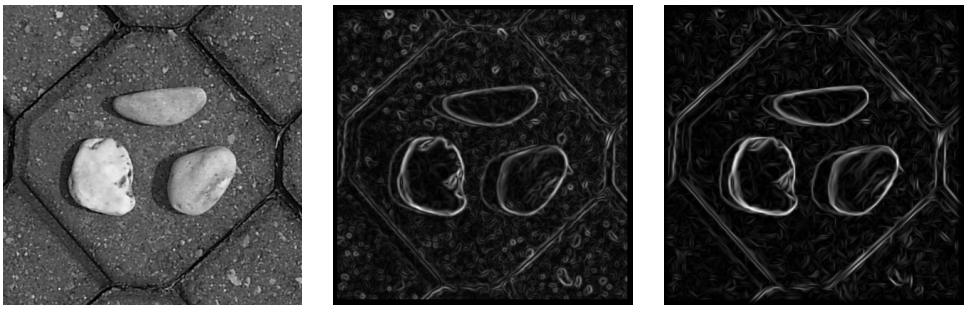
3.2 Anisotropic Diffusion

An important advantage of our problem formulation in (2) is the fact that we are free to choose the weighting factors w_{ij} without loosing existence or uniqueness of the solution. To achieve an anisotropic diffusion behaviour, we use the following definition for the local support area. The orientation angle associated with a pixel is derived by applying an orientation selective Gaussian based filterbank to the input image. Due to the given orientation angle of a pixel the coordinates of its eight-point neighborhood are rotated into the coordinates (x,y) and the corresponding weighting factors w_{ij} are determined from the following function

$$f(x,y) = \begin{cases} 0 & \text{for } y > 0.7 \\ 1.0 + \cos\left(\frac{\pi \cdot y}{0.7}\right) & \text{else} \end{cases} \quad (11)$$

Eq. (11) ensures a strict diffusion along one direction. For a more generous behaviour, the constants 0 and 0.7 in eq. (11) have to be changed. It is clear that $Q_i = \frac{1}{2}\sum_k (w_{ik} + w_{ki}) < Q$ in eq. (9) and hence eq. (11) is well posed.

Fig. 2 demonstrates the anisotropic case on the Lena image. This time, 20% uncorrelated noise is



(a) (b) (c)

Figure 3: Stones image: (a) Input image, (b) gradient magnitude $|\nabla g_x|$ before and (c) after self-organization of the edge map.

added to each pixel of the input image. It becomes obvious that for $c \to \infty$, unlike in the linear isotropic case, a local neighborhood can be defined for which the method does not diffuse to a constant value for all image pixels

4 MODIFICATIONS

4.1 Self-Organization of the Edge Map

As it is mentioned above, we use a Gaussian based filterbank to derive the orientation angle we need to determine the weighting factors w_{ij}. Unfortunately gradient based filters are very sensitive to noise. In order to overcome this problem, we apply the anisotropic regularization to the edge map that is generated by the Gaussian based filterbank, i. e. we apply eq. (2) for $d = 2$ to the edge map that is defined by $(\xi_{(x,1)}, \xi_{(x,2)}) = |\nabla g_x|(cos(\alpha_x), sin(\alpha_x))$ with $|\nabla g_x|$ being the gradient magnitude and α_x the associated orientation at image position x.

During the iterative process, eq. (2) is reformulated for each iteration step, since the weighting factors depend on the edge values that are modified. That is, the orientation of an edgel determines the weighting factors, that are used for one iteration step. This step changes the edgel and therefore its orientation. This leads to a self-organization of the edge map, enhancing salient contours. Note that this proceeding no longer guarantees the uniqueness, the iteration converges to a solution that depends on the starting vector. Nevertheless, regarding the numerical implementation in section 2.2 the convergence of the iteration is still given since $\rho(\mathbf{M}) < 1$. Furthermore, the problem formulation for the image itself stays the same and hence keeps the uniqueness property. Fig. 3 shows the effect of a self-organizing edge map. High frequencies in the input image Fig. 3(a) take effect on the gradient-based edge map in Fig. 3(b). Applying the self-organizing scheme emphasizes "important" edges. This is also of interest for the topic of salient contour inference, see (Mahamud et al., 2003) for example.

4.2 Thresholding the Weighting Factor

To achieve an enhancement of discontinuities along edges the weighting factors can be set to zero for neighboring image elements i, j with $|\xi_{(i,k)} - \xi_{(j,k)}| < s$ for some threshold s. Note that again uniqueness is lost and the result is highly dependent on the starting vector ξ^0. A good choice for ξ^0 is zero or a smoothed version of the input image. For $\xi^0 = \xi_0$ and a small threshold merely no effect will take place even for a high smoothing factor.

5 EVALUATION

5.1 Qualitative Evaluation

The images in Fig. 4(a)-(d) are introduced in (Weickert, 1996) as an example for the denoising properties of different diffusion filters. 70% of the input image in Fig. 4(a) are degraded by noise. Fig. 4(b) shows the linear and (c) the nonlinear isotropic case while Fig. 4(d) is computed with the proposed edge enhancing anisotropic diffusion approach.
In Fig. 4(f) the orientational approvement of the self-organizing process we suggested in section 4.1 can be seen compared to the initial values in Fig. 4(e) that are received from an orientation selective, gaussian based filter. Fig. 4(g) shows the solution of the regularization term proposed in section 2.1 with anisotropic weighting factors derived from the self-organized edge map. Fig. 4(h) shows the same setting, but with thresholded weighting factors. To achieve

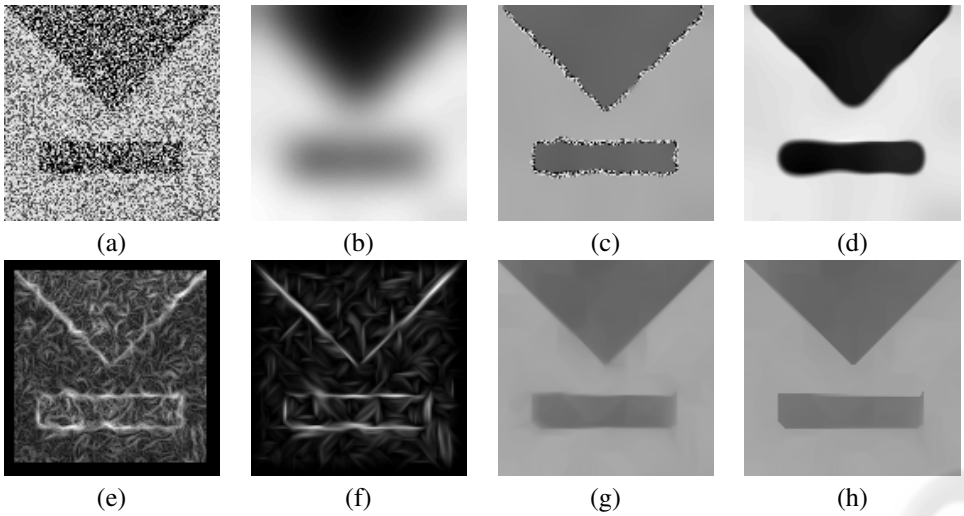


Figure 4: Images (a)-(d) taken from (Weickert, 1996) and (Jähne, 2002) respectively: (a) Input image, (b) linear diffusion, (c) nonlinear isotropic diffusion and (d) nonlinear anisotropic diffusion. Images (e)-(h) demonstrate our approach: (e) Gradient magnitude, (f) gradient magnitude after self-organization of the edge map, (g) result for the anisotropic diffusion proposed in section 3.2 and (h) result with thresholded weighting factors.

Figure 5: Mandrill image: (a) N - input image with 10% uncorrelated noise and (b) R_{250_n} - anisotropic diffusion with $c = 250$.

maximal diffusion, the scale parameter in both cases was set to $c = 1000$. As can be seen in the results our algorithm shows a good noise elimination property with a significant improvement in edge preserving.

5.2 Quantitative Evaluation

The results presented in table 1 are computed on the mandrill image in Fig. 5(a) with 10 percent noise added to each pixel of the input image. To compare two images we take the percentage of pixels with an absolute difference less than a threshold s compared to the total number of pixels $m \cdot d$. For a smoothness factor c the result computed on the noisy input image N is given by R_{c_n}. Table 1 shows the similarity between R_{c_n} and the original image O.

Table 1: Mandrill image: similarity between noisy image (N), original image (O) and computed images (R_c).

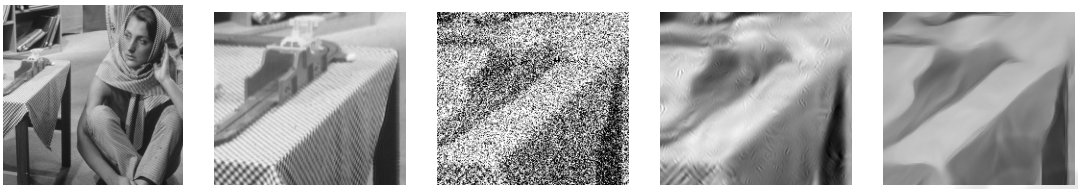
Similarity	$s = 1$	$s = 6$	$s = 16$	$s = 21$
$N \wedge O$	6.19	36.38	79.41	90.36
$R_{0.1_n} \wedge O$	4.04	40.88	85.60	94.40
$R_{0.3_n} \wedge O$	3.99	41.13	84.40	92.75
$R_{0.5_n} \wedge O$	3.90	40.24	81.75	90.29
$R_{0.1_l} \wedge O$	3.90	40.16	83.14	91.92
$R_{250_n} \wedge R_{250_o}$	29.18	98.83	99.97	99.99

Due to the high frequent image content in the original image the diffusion process yields high similarity values only for small smoothness factors. These results show a higher similarity to the original image. than the noisy input image does. Of course, they also show a higher similarity than an linear isotropic filtered image ($R_{0.1_l}$). On the other hand it can be seen that for a large scale the diffusion process comes to similar results when the original image is taken as the input image (R_{250_o}) as well as when the noisy image is taken as the input image (R_{250_n}). This shows that unoriented noise as well as high frequent unoriented image information is eliminated from the input image.

To our knowledge, the noise removal method proposed in (Portilla et al., 2003) and (Guerrero-Colon and Portilla, 2005) respectively is the most accurate one. We therefore compare our results to those given in (Portilla et al., 2003). Here, the denoising performance is measured with the peak signal to noise ra-

(a) (b) (c) (d)

Figure 6: Images used for evaluation in table 2: (a) Lena, (b) Barbara, (c) Boats and (d) House. The input images are degraded with additive Gaussian white noise. All images are taken from *http://decsai.ugr.es/ javier/denoise/*



(a) (b) (c) (d) (e)

Figure 7: Barbara image: (a) original image, (b) magnification of the table region, (c) part of noisy input image with $\sigma = 100$ and PSNR = 8.13, (d) denoising result of Portilla et al. (PSNR = 22.6) and (e) denoising result of our method (PSNR = 22.1).

Table 2: Comparison of our denoising results with the results of Portilla et al., the results are expressed as peak signal to noise ratio (PSNR). The input images, which are shown in Fig. 6, are contaminated with additive Gaussian white noise ($\sigma = 100$). The parameters of our method are tuned to show best performance on the Lena image. All images are computed with the same parameters.

$\sigma = 100$	Lena	Barbara	Boats	House
Portilla et al.	25.64	22.61	23.75	25.11
Our method	25.65	22.07	23.13	24.92

tio PSNR $= 20\log_{10}(255/\sigma_e)$ with σ_e being the error standard deviation. Looking at the PSNR values our method reveals a notable disadvatage. The smoothing, especially for large smoothness factors, leads to a shrinkage of the histogram. For the sake of comparability, we therefore applied a linear histogram equalization on our results, but, as can be seen in Fig. 7(e), there still remains a systematic deviation of a regions greyscale level compared to the original image (Fig. 7(b)). Our results are computed with a smoothing factor $c = 4.0$ and 60 iteration steps. The starting vector is the noisy input image and in a preprocessing step, we perform the self-organization of the edge map as it is described in section 4.1. For the self-organization, we used a smoothing factor $c = 8.0$. As can be seen in table 2, except for the Lena image, our approach is outperformed by the method proposed in (Portilla et al., 2003). Nevertheless it can be observed that for high noise levels, our method produces less artifacts, which can be a crucial issue, not only for subsequent image processing steps, but also for the viewers subjective appraisement of the image quality.

6 MORE RESULTS

Anisotropic diffusion processes also feature artistic aspects. In (Weickert, 1996) , paintings of van Gogh were used for coherence enhancing anisotropic diffusion resulting in images that yield very different impressions compared to the original paintings. In Fig. 8(b) the anisotropic regularization we propose in section 3.2 is used to eliminate the pointillism in a painting of Georges Seurat (Fig. 8(a)). In Fig. 8(c) and (d) the isotropic case in section 3.1 is combined with the thresholding in section 4 and an oversmoothing of the images in order to alienate the original images.

7 CONCLUSIONS AND OUTLOOK

We presented a novel regularization-based approach to anisotropic diffusion that provides a clear mathematical formulation of the regularized diffusion problem and its solution. The solution is unique and the iteration process to derive this solution is shown to be convergent for any scale, i. e. smoothness factor.

(a) (b) (c) (d)

Figure 8: (a) Un dimanche après-midi à l'île de la grande jatte, 1884 (b) Result after applying anisotropic diffusion (c) Street scene in Florence (d) Somewhere in Ireland.

The capability of our approach was demonstrated on a variety of greyscale and color images. By our approach we hope to have closed a gap between classic diffusion filters and regularization methods. Future examinations will include verifying the rotation invariance properties of the local support area proposed in section 3.2. Future work also will address the effect of histogram shrinkage mentioned above and the improvement of the comparability of the noise removal results.

REFERENCES

Aubert, G., Barlaud, M., Blanc-Feraud, L., and Charbonnier, P. (1997). Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Imag. Process.* 5(12).

Brockers, R., Hund, M., and Mertsching, B. (2005). Stereo Matching with Occlusion Detection Using Cost Relaxation. In *IEEE International Conference on Image Processing (ICIP 2005)*, volume III, pages 389 – 392.

Guerrero-Colon, J. A. and Portilla, J. (2005). Two-level adaptive denoising using gaussian scale mixtures in overcomplete oriented pyramids. IEEE International Conference on Image Processing (ICIP 2005), vol. I, 105-108.

Hund, M. (2002). Disparitätsbestimmung aus Stereobildern auf der Basis von Kostenfunktionen. Diploma thesis, University of Paderborn.

Ito, K. and Kunisch, K. (1999). An active set strategy based on the augmented lagrangian formulation for image restoration. M2AN Math. Model. Numer. Anal., Vol. 33, pp. 1-21.

Jähne, B. (2002). *Digital image processing (5th ed.): concepts, algorithms, and scientific applications*. Springer-Verlag.

Mahamud, S., Williams, L. R., Thornber, K. K., and Xu, K. (2003). Segmentation of multiple salient closed contours from real images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(4).

Nordstrom, K. N. (1989). Biased anisotropic diffusion–a unified regularization and diffusion approach to edge detection. Technical Report UCB/CSD-89-514, EECS Department, University of California, Berkeley.

Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639.

Portilla, J., Strela, V., Wainwright, M., and Simoncelli, E. (2003). Image denoising using scale mixtures of gaussians in the wavelet domain. IEEE Trans. Image Proc., In Press. 2003.

Scharr, H. and Weickert, J. (2000). An anisotropic diffusion algorithm with optimized rotation invariance. In G. Sommer, N. Kr uger, C. Perwass (Eds.), Mustererkennung.

Weickert, J. (1996). Anisotropic diffusion in image processing. Ph.D. thesis, Dept. of Mathematics, University of Kaiserslautern, Germany.

Weickert, J. (1999). Coherence-enhancing diffusion of colour images. *Image Vision Comput.*, 17(3-4):201–212.