

Introduction of the mean shift algorithm in SAR imagery: Application to shadow extraction for building reconstruction

François Cellier¹, H el ene Oriot¹, Jean-Marie Nicolas²

¹ONERA - BP72 - 29 avenue de la Division Leclerc 92322 Ch atillon, France ; cellier@onera.fr ; helene.oriot@onera.fr

²ENST - 46, rue Barrault - 75013 Paris, France ; nicolas@tsi.enst.fr

Keywords: Mean shift, SAR imagery, shadow extraction, filtering

ABSTRACT: The mean shift is a non-parametric algorithm that shifts each sample to the mode of the statistic it belongs to. This filter was originally dedicated to cluster analysis of data and has been recently extended to image processing. In this paper, we are introducing the mean shift algorithm in SAR imagery. It is shown that, after an adapted preprocessing step to reduce the granular aspect of amplitude SAR images, the mean shift gives interesting performances compared to usual SAR parametric filters. This is achieved without any assumptions concerning the underlying statistics involved in the images. An extension of this filter to the coherence image is then proposed. Amplitude and coherence images are simultaneously filtered. It is shown that these two filtered images can be used to improve shadow extraction for building reconstruction.

1 INTRODUCTION

The reconstruction of buildings in urban environment using high-resolution SAR imagery is a challenging field of investigation. Indeed, with the recent technology improvements, it becomes possible to use SAR images taken from airborne sensors for recognition tasks. Nevertheless, the SAR phenomenology makes the scene understanding very difficult and the need of semi or fully automatic tools to help interpreters is real. In this paper, we aim at extracting shadows from high-resolution interferometric SAR (*X-Band*) images in urban environment, from the RAMSES sensor (*ONERA, France*), using both amplitude and coherence information. However shadows are not easy to extract and a filtering step is needed. Several parametric filters using assumptions on the statistics of the SAR images have been proposed (for instance, the Lee filter (Lee 1980), the Gamma MAP filter (Kuan *et al.* 1987), the Fisher MAP (Nicolas 2003)). In this paper, we are considering a non parametric approach by using the mean shift algorithm (Fukanaga & Hostetler 1975).

In section 2, we present some basics on non-parametric density estimation used in the mean shift algorithm. Then, we detail the original mean shift algorithm and its recent extension to image processing (Commaniciu & Meer 2002). The next section shows the behaviour of the mean shift filter and the influence of an adapted preprocessing step on synthetic images. We also compare it with some usual parametric filters used in SAR imagery. In the fourth section, results are shown on images from RAMSES, a French sensor of the ONERA. Then, we extend the mean shift filter to the coherence image where shadows are also visible and we propose a fusion algorithm of the amplitude and coherence filtered images.

2 THE MEAN SHIFT ALGORITHM

In this section we briefly describe the non parametric density estimation also called density smoothing and the principle of the mean shift algorithm.

2.1 Non parametric density estimation

Let f be a probability density function, χ a random variable associated to f and $\Omega = \{x_1, \dots, x_n\}$ a realisation drawn from χ . A convolutive kernel is a bilinear form defined by:

$$K : \begin{cases} \Omega \times \Omega \rightarrow \mathfrak{R} \\ (x_i, x_j) \rightarrow K(x_i - x_j) \end{cases}$$

For any symmetric kernel K , the associate profile k is defined by $K(x) = c.k(\|x\|^2)$, where c is a normalising parameter. Two usual symmetric kernels are the truncated gaussian (K_G) and the Epanechnikov (K_E) kernels defined by (Silverman 1986):

$$K_G(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\|x\|^2}{2}\right) \quad K_E(x) = \frac{1}{2}(1 - \|x\|^2)$$

Both kernels have a finite support, defined by the bandwidth h : $K\left(\frac{x}{h}\right) = 0$ if $\|x\| > h$.

The Parzen estimator of f at the sample x_j using K is defined as:

$$\hat{f}_{h,K}(x_j) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_j - x_i}{h}\right),$$

The Parzen estimator is said to be non-parametric as it makes no assumption about the statistics that are estimated. The shape and the number of modes of the resulting density depend on the choice of the bandwidth parameter (Silverman 1986).

2.2 Principle of the mean shift algorithm

The mean shift algorithm is introduced in Fukunaga & Hostetler 1975 as a non-parametric clustering algorithm. It works in an iterative way to shift each sample uphill toward the local mode of the smoothed density. The convergence of the algorithm is proven for the Gaussian kernel (Cheng 1985) and for any kernel with a convex and monotonically decreasing profile, such as the Epanechnikov kernel (Commaniciu & Meer 2002). We consider in this paper only kernels satisfying these hypotheses.

Be $\{x_1^0, x_2^0, \dots, x_n^0\}$ the samples at the iteration 0. We note $\{x_1^m, x_2^m, \dots, x_n^m\}$ the samples obtained at iteration m .

For all samples, the smoothed density is given by :

$$\hat{f}_{h,K}(x_j^m) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x_j^m - x_i^m}{h}\right) = \frac{1}{n} \sum_{i=1}^n \frac{c}{h} k\left(\left\|\frac{x_j^m - x_i^m}{h}\right\|^2\right), j=1 \dots n$$

It is shown in Cheng 1985 that, for each sample the mean shift term at sample x_j^m , noted $ms_{h,K}(x_j^m)$, can be linked to the gradient of the smoothed density : $ms_{h,K}(x_j^m) = \alpha \cdot \hat{\nabla} f_{h,K}(x_j^m)$, where $\alpha > 0$.

By updating the data with $x_j^{m+1} = x_j^m + ms_{h,K}(x_j^m)$, each sample is iteratively shifted to the local mode of the smoothed density. In case of unimodal statistics, all samples following the same statistic converge to the same mode, which means that the mean shift is a clustering algorithm. The algorithm stops at iteration l when all the samples reach their respective mode, that is to say $ms_{h,K}(x_j^l) = 0$ for $j=1 \dots n$.

2.3 Extension of the algorithm to the spatial domain for image processing

The mean shift algorithm is extended to image processing by taking into account the spatial neighbourhood of the samples (*pixels*) in Commaniciu & Meer 2002. The authors introduce the joint domain : each sample is now described both by its range domain (gray level, RGB components...) and its spatial domain (position in the image).

$\forall i \in [1 \dots n]$, $z_i^m = \{s_i^m, a_i^m\}$, where s represents the spatial information and a the range information. Then, the kernel is defined by the product of two kernels, K_A for the range domain and K_S for the spatial domain.

The smoothed density is written as:

$$\hat{f}(z_j^m) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_A h_S} K_A \left(\frac{a_j^m - a_i^m}{h_A} \right) K_S \left(\frac{s_j^m - s_i^m}{h_S} \right)$$

where h_A is the bandwidth in the range domain and h_S the one in the spatial domain. Only the samples lying within h_A and h_S of z_j^m , will contribute to the estimation of the density at z_j^m . The update of the samples is made by: $a_j^{m+1} = a_j^m + ms_{h_A, K_A}(a_j^m)$ and $s_j^{m+1} = s_j^m + ms_{h_S, K_S}(s_j^m)$.

The convergence occurs simultaneously in both domains and is reached at iteration l when $ms_{h_A, K_A}(a_j^l) = 0$ and $ms_{h_S, K_S}(s_j^l)$ for $j=1 \dots n$.

The clustering nature of this algorithm appears in both domains. In the same way as previously, in the range domain, samples belonging to the same statistic converge to the same mode. Moreover, by updating the data in the spatial domain, the localisation of the samples at each iteration is likely to change. Samples following the same amplitude statistic spatially gather in basin of attraction.

3 EVALUATION OF THE MEAN SHIFT FILTER ON SYNTHETIC SPECKLED IMAGES

In order to evaluate the mean shift algorithm for SAR imagery, we compare its performances to the performances obtained with parametric and non-parametric filters. In this section we present the evaluation method as well as results on synthetic speckled images. Results on real images are presented in section 4.

3.1 Evaluation method

We generate a synthetic data set following the Rayleigh-Nakagami distribution, also called Generalised Rayleigh distribution. This data set consists in images divided into two zones of 128x256 pixels. By adapting the means of each area, we obtain 14 contrasts from 7.78dB to 0.41dB. For each contrast, 100 single look images are considered. An example of these images is shown on figure 1a.

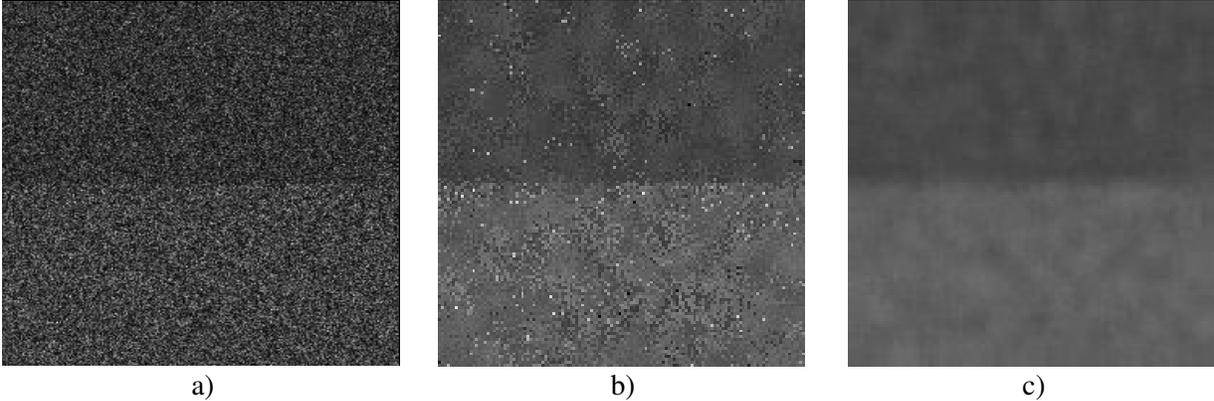


Figure 1 – a) synthetic image (contrast 1.76dB), multilook image filtered with
b) an adapted bandwidth, c) a too large bandwidth

To compare the different filters in a quantitative way, we use the receiving operating curve (ROC) which consists in plotting the probability of detection (p_d) with respect to the probability of false alarm (p_{fa}). To do that, we must define those probabilities. We note A the whole image, D the extracted region of the filtered image whose amplitude is lower than a threshold th , and V the true - darkest half of the synthetic image.

We define the probabilities of detection and false alarm as:

$$p_d(th) = \frac{D(th) \cap V}{V} \quad \text{and} \quad p_{fa}(th) = \frac{D(th) - D(th) \cap V}{A - V}$$

By using different thresholds from the minimum to the maximum values of A , we obtain several couples $(p_{fa}(th), p_d(th))$ that represents the ROC of the filter on image A .

3.2 Choice of the bandwidth parameters

We test here the influence of the spatial and range (amplitude) bandwidths, respectively noted h_S and h_A , on the performance of the mean shift filter. The two areas of the synthetic images can be regarded as homogenous areas, which means that the size of the considered area is large compared to the filtering window. All the results on synthetic images are shown for a contrast of 1,76dB and are similar for other contrasts. The influence of the range bandwidth is shown on figure 2a.

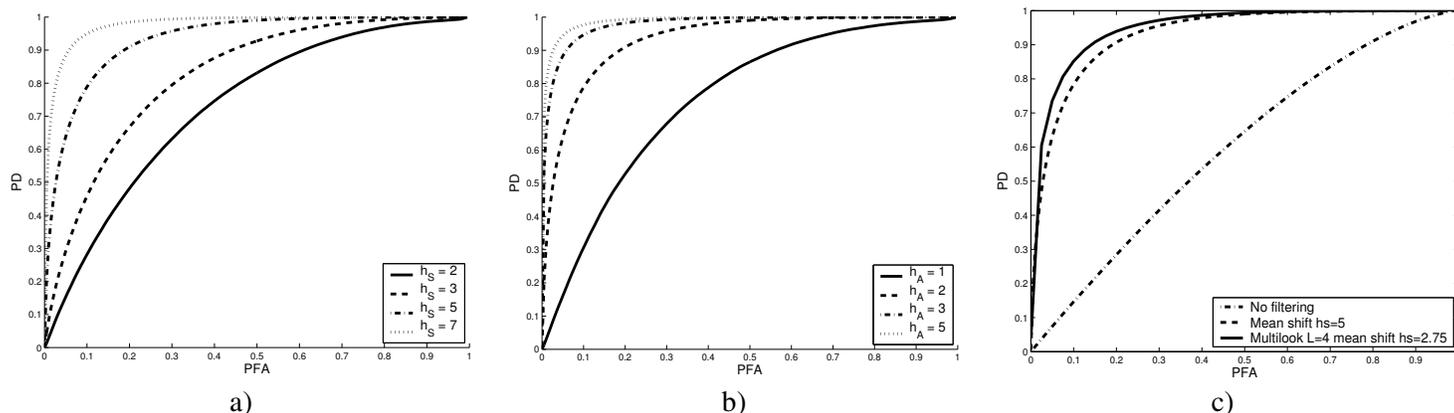


Figure 2 – Influence of a) the range (amplitude) bandwidth, b) the spatial bandwidth and c) the multilook

Taking a larger bandwidth increases the efficiency of the filtering because it enables to consider more samples belonging to the same statistic. However, using a too large bandwidth blurs the border between the two areas (cf. figure 1c). On figure 2b, we can see that using a larger spatial bandwidth improves the results of the filtering. Once again, considering more pixels helps to obtain a better smoothing of the density. On the other side, the computational time drastically increases.

3.3 Influence of the multilook

As the mean shift algorithm is based on histogram smoothing, we check the influence of the multilook, used as a preprocessing step. The multilook is a very common process in SAR imagery which enables to reduce the granular aspect of SAR images resulting in less noisy local histograms. It is usually done by considering the image in blocks of L pixels and taking the non-coherent mean of the L pixel of the intensity image.

To evaluate the mean shift in single and multi look, we need to consider the same neighbourhood. The mean shift filtering window is a disc which diameter can be linked to the bandwidth by $d = 2h_S - 1$. Considering an

L -look preprocessing using square blocks, the diameter of the filtering window in multilook should be $d.L^{-1/2}$. Choosing a spatial bandwidth of 5 in single look gives a bandwidth of 2.75 in multilook with $L=4$. As expected, the combined use of the preprocessing step improves the efficiency of the mean shift (see figure 2c) by reducing the noise in the histogram.

3.4 Comparison with usual SAR parametric filters

Here, we are comparing the mean shift with several filters commonly used in SAR imagery. To take into account the specific nature of the SAR images, the chosen filters are parametric, that is to say, they use statistical models to estimate the underlying reflectivity of the pixels in the image.

Those filters are the Lee (Lee 1980), the Kuan (Kuan *et al.* 1985) and the Frost (Frost *et al.* 1982) which are enhanced in (Lopes *et al.* 1990). They are based on the minimisation of the mean square error (MMSE) which assumes that the true reflectivity of a pixel is a linear combination of the value of the pixel and its neighbourhood. They also assume that the model fitting the SAR image is the Rayleigh-Nakagami distribution.

We also choose the Gamma MAP (Kuan *et al.* 1987) and Fisher MAP (Nicolas 2003a) filters. They are based on the maximum *à posteriori* (MAP) criterion. The assumed models are respectively the K and Fisher distributions.

For a better understanding of the figures, we only show the results with the best MMSE filter (the Frost filter in this case) and the best MAP filter (Gamma MAP filter in this case) using a 9x9 window. On figure 3, the ROCs of the mean shift filter and the parametric filters are shown on multilook images ($L=4$, $h_S=5$).

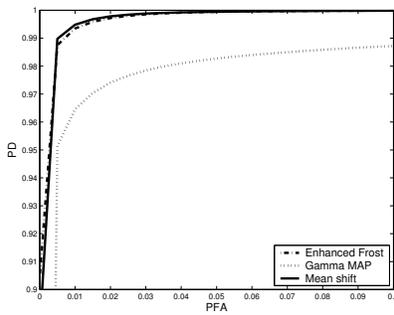


Figure 3 – Comparison of parametric filters with the mean shift on multilook images ($L=4$)

We can see that the Frost filter and the mean shift filter give the best results. It is important to notice that, in this study, synthetic images correspond to fully developed speckle without any other noise (thermal noise or outliers). The mean shift, which makes no assumption on the statistic of the image, gives similar results as the best parametric filter used in an ideal case (Rayleigh-Nakagami distribution). This tends to prove that the mean shift filter is quite robust. In the next section we study the performances of the mean shift filter on real images.

4 EVALUATION OF THE MEAN SHIFT ON RAMSES IMAGES

RAMSES is a French SAR airborne sensor belonging to ONERA (Boutry 1998). In this experiment we use high resolution interferometric data in X band. An amplitude image, where shadows are contrasted, is shown on figure 5a.

We use the mean shift algorithm as well as the parametric filters presented in section 3.4 on this amplitude image in order to extract shadow areas. We plot, on figure 5b, the ROC curves of the best MMSE and MAP filters which are respectively the Frost and the Fisher MAP filters and compare them to the mean shift curve. As we can see, for a p_{fa} higher than 1%, the mean shift is more efficient than the parametric filters. We show, on figure 5d-e, the extracted shadows in black and the false alarms in white for the mean shift and Fisher MAP filter at a p_{fa} of 2%. To delineate the shadow ground truth we did not consider regions where triple bounced scatterers appear (Dupuis *et al.* 2000). Shadow pixels of the ground truth that are not extracted are represented in light gray. We can see that the main shadow is well extracted on the mean shift filtered image (figure 5d). With both filters, the small shadow is poorly extracted because it has a higher backscattering level.

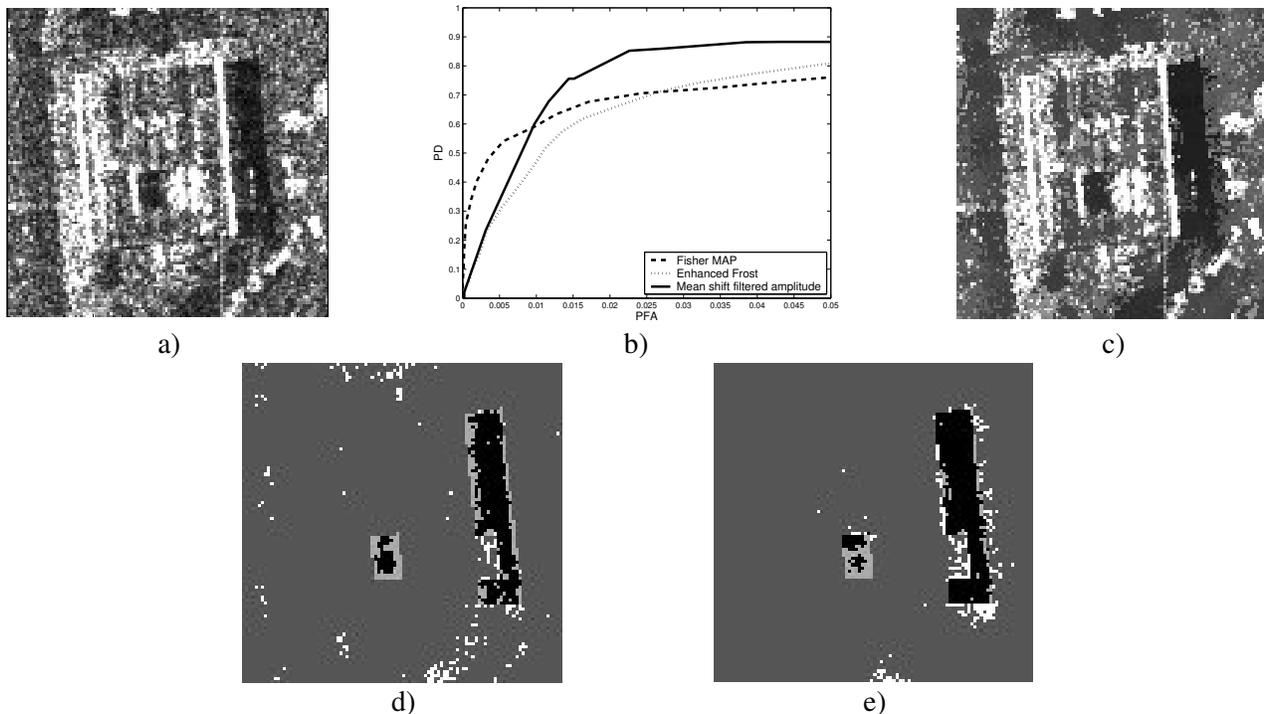


Figure 5 – a) amplitude, b) comparison of the mean shift with parametric filters, c) amplitude filtered by the mean shift, shadows extracted on the d) Fisher MAP, e) mean shift filtered images. Extracted shadow pixels, false alarms and missed shadow pixels are respectively in black, white and light gray.

The shadows are visible on both amplitude and coherence images (see figure 6a-b). Therefore, we extend the mean shift to take into account the coherence image.

5.1 Extension of the joint domain to coherence

In a similar way as in §2.3, we introduce an extension of the mean shift filter to the coherence image.

The estimated density is: $\hat{f}(z_j^m) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_A h_S} K_A \left(\frac{a_j^m - a_i^m}{h_A} \right) K_C \left(\frac{c_j^m - c_i^m}{h_C} \right) K_S \left(\frac{s_j^m - s_i^m}{h_S} \right)$, where K_C and h_C are

the kernel and the bandwidth in the coherence domain.

The theoretical coherence Γ is computed using the following equation (Maître 2001):

$$\Gamma = \left\| \frac{E(z_1 z_2^*)}{\sqrt{E(|z_1|^2) E(|z_2|^2)}} \right\|, \text{ where } z_1 \text{ and } z_2 \text{ are the two images of the interferometric couple.}$$

Γ is, by definition, between 0 and 1. Practically the coherence image (empirical coherence) is computed on a finite number of pixels in a similar way as the multilook image, that is to say, by considering the images in blocs of L pixels:

$$\gamma = \left\| \frac{\sum_{k=1}^L z_{1,k} z_{2,k}^*}{\sqrt{\left(\sum_{k=1}^L z_{1,k} z_{1,k}^* \right) \left(\sum_{k=1}^L z_{2,k} z_{2,k}^* \right)}} \right\|$$

The statistical distribution of the coherence is, for $L \geq 2$ (Maitre 2001):

$p(\gamma|\Gamma, L) = 2(L-1)(1-\Gamma^2)^L \gamma(1-\gamma^2)^{L-2} {}_2F_1(L, L; 1; \gamma^2 \Gamma^2)$, where ${}_2F_1$ is the hypergeometric function. In the shadows, only noise is measured because no signal is backscattered to the sensor. Therefore the theoretical coherence is zero. The distribution of the empirical coherence in shadows is: $p(\gamma|\Gamma = 0, L) = 2(L-1)\gamma(1-\gamma^2)^{L-2}$.

In order not to lose too much in resolution, we restrain ourselves to use 4-look images. On figure 4, we plot the distribution of the empirical coherence for $L=4$. As we can see, the empirical coherence goes from 0 to 1. Therefore we choose the bandwidth in the coherence domain equal to 1.

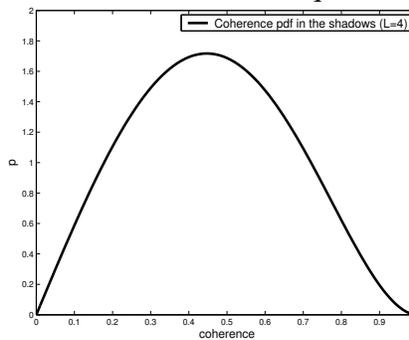


Figure 4 – Distribution of the empirical coherence in shadows ($L=4$)

For the sample z_j^m , the density smoothing occurs in both amplitude and coherence domains by considering only samples that are lying within the amplitude, coherence and spatial bandwidths. A coherence bandwidth equal to 1 implies that all samples in the coherence domain lay within the coherence bandwidth of z_j^m . Therefore the coherence is filtered on samples which are lying within the amplitude and spatial bandwidths of z_j^m .

In the same way as previously, the global mean shift is composed of three mean shift terms. From this filtering result two filtered images (amplitude and coherence) which need to be fused in order to have a common shadow extraction.

5.2 Fusion algorithm

Shadows can be characterised by low amplitude and low coherence. To extract pixels satisfying both conditions we use the following fusion algorithm. By thresholding regularly the filtered coherence and amplitude images from the minimum to the maximum, we obtain two sets of hypothetical shadows. The thresholds are noted $S_C = \{s_C\} = \{0, \dots, step_C, \dots, 1\}$ for the coherence and $S_A = \{s_A\} = \{0, \dots, step_A, \dots, \max(A)\}$ for the amplitude. We note respectively $C(s_C)$ and $A(s_A)$ the coherence and amplitude images thresholded at s_C and s_A . For each s_C , we find the threshold on the amplitude image that maximises the following criterion:

$$d = \frac{C(s_C) \cap A(s_A)}{C(s_C) \cup A(s_A)}$$

That is for each coherence threshold, we find the amplitude threshold that maximises the

percentage of pixels which are labeled as shadow on both images. We then obtain, for each coherence threshold, an extracted shadow area that is the intersection of the two shadow areas.

5.3 Results

On figure 6c, we present the ROCs for the parametric filters, the extended mean shift and the fusion algorithm. As we can see, the mean shift, either on amplitude or coherence images, has better performances than the parametric filters. This is also visible on the extracted shadows at a p_{fa} of 2% on the amplitude image (figure 6g). Moreover, we can see some complementarity between shadows extracted from amplitude and coherence (figure 6h) images. For instance, the shadow on the left of the image presents not only some part appearing in one of the two images but also another part, on the top, that is not extracted on both images. The fusion algorithm enables a fast complete extraction of this shadow (figure 6i). The extraction of the other shadows is also improved.

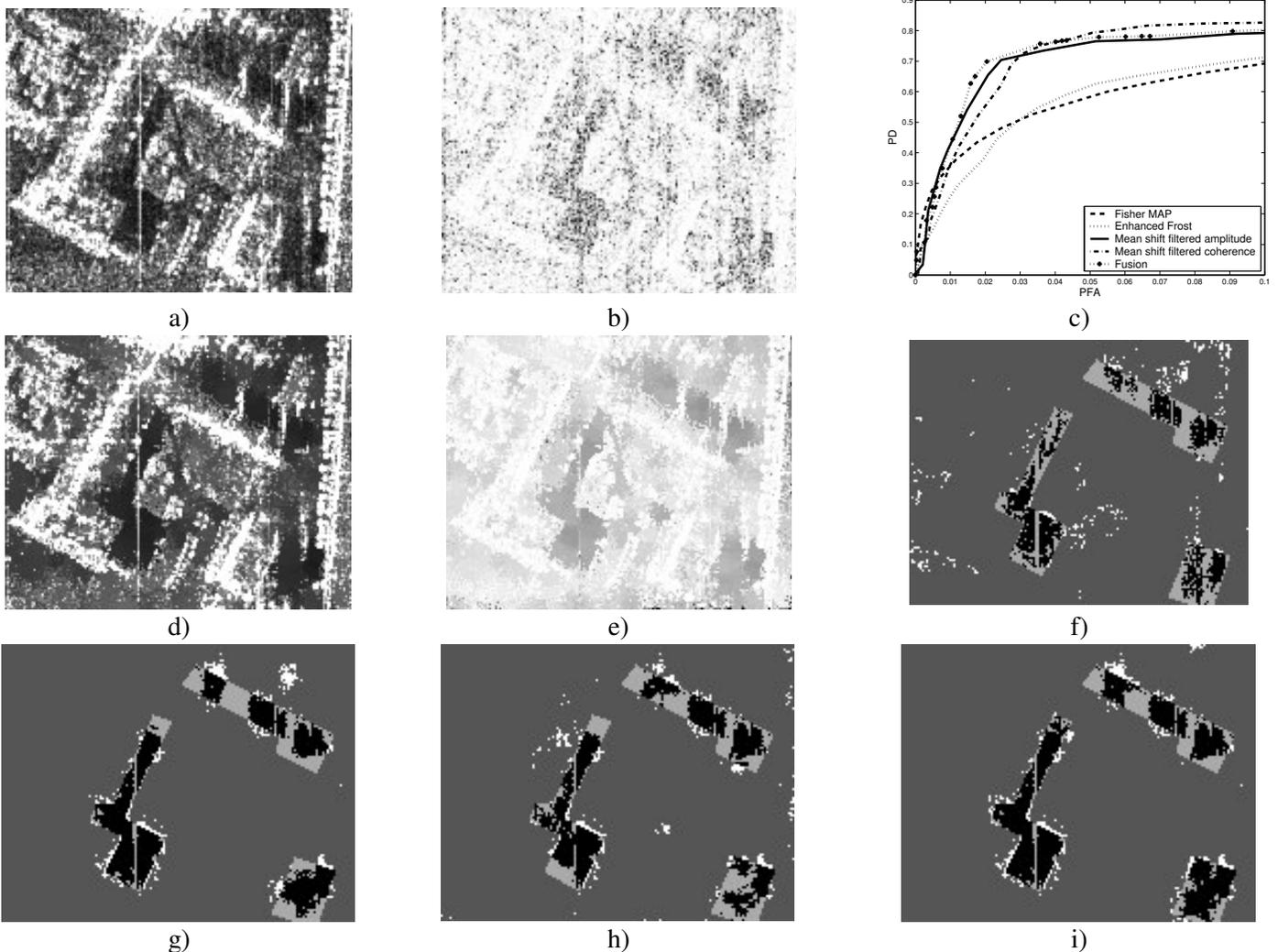


Figure 6 – a) amplitude, b) coherence images, c) comparison of the extended mean shift with the parametric filters, d) amplitude and e) coherence filtered by the mean shift, shadows extracted on amplitude image filtered by the f) Fisher MAP, g) mean shift, h) shadows extracted on coherence image filtered by the mean shift, i) shadows extracted on the fused image.

Extracted shadow pixels, false alarms and missed shadow pixels are respectively in black, white and light gray.

6 CONCLUSION

We introduced in this paper the mean shift algorithm in SAR imagery. We have seen that, the performances of the mean shift, which requires no assumptions on the statistics, are similar to those of some parametric filters used in SAR imagery on synthetic images following a Rayleigh-Nakagami distribution with ($L=4$). In addition, interesting results have been shown on RAMSES amplitude images. We further extended this filter to the shadow extraction by considering the coherence image. By using a fusion algorithm of the amplitude and coherence filtered images, we improved the shadow extraction.

7 REFERENCE

- Boutry J.M., 1998, Image quality parameters of a high resolution experimental SAR, ONERA -PUBLICATIONS- TP 1998, Vol 204
- Cheng, Y., 1995, Mean shift, mode seeking and clustering, IEEE Transactions on pattern analysis and machine intelligence, Vol 17, n°8, pp790-799.
- Comaniciu, D. & Meer, P., 2002, Mean shift : a robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol24, n°5, pp603-619.
- Dupuis X, Dupas J, Oriot H, 3D Extraction From Interferometric High Resolution SAR Images Using the Ramses Sensor, Proceedings of EUSAR'2000, Munich, may 2000, pp505-508.
- Frost, V.S., Stiles, J.A., Shanmugan, K.S., Holtzman, J.C., 1982, A model for radar images and its application to adaptive digital filtering of multiplicative noise, IEEE Transaction on pattern analysis and machine intelligence Vol 4, n°2, pp157-166.
- Fukanaga, K. & Hostetler, L., 1975, The estimation of the gradient of a density function, with applications in pattern recognition, IEEE Transactions on information theory, Vol 21, n°1, pp32-40.
- Kuan, D.T., Sawchuk, A., Strand, T.C., Chavel, P., 1985 Adaptive noise smoothing filter for images with signal dependant noise. IEEE Transactions on pattern analysis and machine intelligence, Vol 7 n°2, pp165-177.
- Kuan, D.T., Sawchuk, A., Strand, T.C., Chavel, P., 1987, Adaptive restauration of images with speckle. IEEE Transactions on acoustics, speech and signal processing, Vol 35, n°3, pp373-383.
- Lee, J.S., 1980, Digital image enhancement and noise filtering by use of local statistics, IEEE Transaction for pattern analysis and machine intelligence, Vol 2, pp165-168.
- Lopès, A., Touzi, R. & Nezry, E., 1990, Adaptive speckle filters and scene heterogeneity, IEEE Transactions on Geoscience and Remote Sensing, Vol 28, n°6, pp992-1000.
- Maître, H., 2001, Traitement des images de RSO, Ed Hermes
- Nicolas, J.M., 2003, A Fisher-MAP filter for SAR image processing IGARSS'03, Toulouse, pp1996-1998
- Silverman, B.W., 1986, Density estimation for statistics and data analysis, Ed. Chapman & Hall
- Tison, C., Nicolas, J.M., Tupin, F., 2003, Accuracy of Fisher distributions and log-moment estimation to describe amplitude distributions of high resolution SAR images over urban areas, IGARSS'03, Vol 3, 1999-2001