Adaptive Mixtures of GSMs for Image Denoising

A project report

submitted by

Mutnuru Sarath Kumar

in the partial fulfilment of the requirements

for the award of the degree of

Bachelor of Technology

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUE OF TECHNOLOGY-MADRAS

THESIS CERTIFICATE

This is to certify that the thesis titled **Adpative Mixtures of GSMs for Image Denoising**, submitted by **Mutnuru Sarath Kumar**, to the **Indian Institute of Technology, Madras**, for the award of the degree of **Bachelor of Technology**, is a bonafide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof Kaushik Mitra
Project Guide
Assistant Professor
Department of Electrical Engineering
IIT-Madras
Chennai-36

ACKNOWLEDGEMENTS

This work would not have been possible without the guidance and the help of several people. I take this opportunity to extend my sincere gratitude to all those who made this thesis possible.

First, I would like to thank all my teachers who bestowed me with good academic knowledge. I am indebted to my advisor Prof. **Kaushik Mitra** whose expertise, generous guidance and support made it possible for me to work on a topic that was of great interest to me. I would also like to thank my lab mate and dear friend Anil Vadathya for sharing his valuable ideas and helping me whenever I am stuck with some problem.

I would like to thank my family for giving support and guidance all through my life. I would also like to thank all my friends and well-wishers for helping me in difficult times and being a good source of support and guidance.

ABSTRACT

Mixture of Conditional Gaussian Scale Mixture(MCGSMs) ,which is a slight modification of GMM is a good image prior and performs well in image processing tasks like compressive sensing,image recovery,image inpainting ,single pixel camera applications.

But when MCGSM is applied in context of image denoising by using it as a prior and MAP framework,the denoised images are found to be smoothening and losing their texture.So our goal here is to try to improve denoising using MCGSM by adapting it to the given noisy data.For this we try approaches like modifying the gradient and training the MCGSM on similar clusters

Chapter 1 Introduction

1.1 Introduction

Image priors are very crucial for various image processing tasks like image denoising,image inpainting,compression and reconstruction and also in image recognition tasks.Image priors are probabilistic models of natural images.The more common goal is to learn image statistics effectively.Image priors are generative type of networks and generative networks are task independent networks,trained once on a image,they can be used for various tasks where as Discriminative networks are more of task specific.

There are undirected and directed models with regard to image priors.Directed models present easy ways of inference,evaluation than undirected models.Other important advantage of directed models is that they allow for effective decomposition of distribution defined.

For instance ,

$$P(X) = \prod_{xy} P(X_{xy}|Xa_{xy}) \tag{1}$$

where $Xa_{x,y}$ refers to the causal neighbourhood of pixel X_{xy}

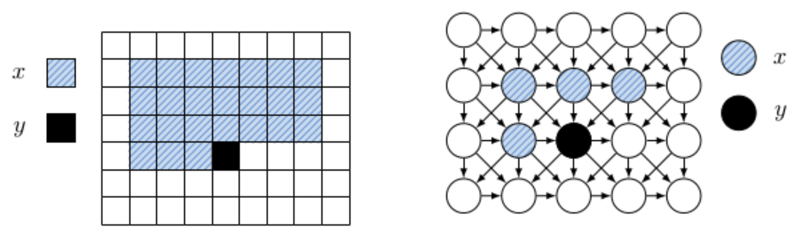
The justification for this points back to primary rules of conditional proability.Consider a joint distribution among four variables x, y, z, w and consider their splitting like the following

$$\begin{aligned} P(x,y,z,w) &= P(x|y,z,w)\,P(y,z,w) && (2)\\ &= P(x|y,z,w)\,P(y|z,w)\,P(z,w) && (3)\\ &= P(x|y,z,w)\,P(y|z,w)\,P(z|w)\,P(w) && (4) \end{aligned}$$

This is similar to what saying as

$$P(\bar{x}) = \prod_i P(x_i | x_{<i}) \tag{5}$$

and we consider $x_{<i}$ as the casual neighbourhood of pixel x_i and $\bar{x}$ being the whole image.Here the point to be noted is that there is independent assumption has been made.This way of modelling the whole image by using a directed model in which pixels are assumed to depend directly only on the ones to the left and above of it.This set is called a *causal neighbourhood* and resulting directed model is *causal random field.*However the pixel still depends on every neighbour,this is a mild modification of markov blanket



(a) 24 pixels in causal neighbourhood (b) four pixels in causal neighbourhood

Figure 1: causal neighbouhood

1.2 Summary

In chapter-2 we introduce the MCGSMs , denoising using them as prior and compare the results with existing techniques and point out its shortcomings.Later, in chapter-3 we present various methods which to adapt the mcgsm and to improve it.

Chapter 2 MCGSM

2.1 Theory

MCGSMs are a generalization of Gaussian Mixture Models but their parametrization makes them suitable for natural image statistics.They model the conditional distribution of a pixel given its causal neighbourhood.Several parametrizations have been suggested in modelling a pixel given its parent pixels,MCGSM is one such parametrization,it conditions a pixel on pixels of its causal neighbourhood.This gives pixels an order and a flexible but tractable parametrization.

Let x be a image patch,x_{ij} be the intensity of the image at pixel location ij .Also,let $x_{<ij}$ indicate the pixels x_{mn} with $m \leq i$ and $n \leq j$ .Then we can introduce a conditional factorization on the whole image patch like below

$$p(x;\theta) = \prod_{ij} p(x_{ij}|x_{<ij};\theta) \tag{6}$$

This distribution does not assume any independence between pixels.It is a straight forward application of chain rule.The parameters of this model are θ.Usually the parameters may differ for different pixels which results in a model like the one below.

$$p\left(x;\left\{\theta_{ij}\right\}\right) = \prod_{ij} p\left(x_{ij}|x_{<ij};\left\{\theta_{ij}\right\}\right) \tag{7}$$

But this would drastically increase the number of paramters in the model.One way of decreasing parameters is associating a Markov assumption and decreasing the neighbourhood and other way is to assume stationarity or shift invariance,in which case ,only one set of $\left\{\theta_{ij}\right\}$ is needed to learnt and can be applied to any pixel of the image.

However to complete the model,distribution of a pixel given its causal neighborhood needs to specified.That means, each one in the product of *eqn* (6) needs to modelled.That is $p(x_{ij}|x_{<ij};\theta)$ needs to modelled.As we assumed stationarity,the problem reduces to specifying a single conditional distribution.

We consider *Gaussian Scale Mixtures(GSMs)* , as they do very good in modelling natural statistics of image.The equation regarding to *GSM is as below.*

$$p(x) = \int \psi(z) N(x;\mu, zC) dz$$

where $N(x;\mu, zC)$ is a normal distribution over x with mean μ and covariance zC and $\psi(z)$ is univaraite density over scales.MCGSM is a conditional form of mixture of *GSMs* ,with each pixel being conditioned on its causal neighbourhood.We use finite mixture of *GSMs* ,meaning we use only a finite number of scales,and each mixture has equal weight for component and scale.

The equations governing MCGSM with reference to the above explanation turn out to be

$$p(y|x) = \sum_{c,s} p(c,s|x) p(y|c,s,x)$$

The equation is similar to the framework of *mixture of experts* with $p(y|c,s,x)$ being as the *expert* and $p(c,s|x)$ as the *gate.*In this type of formulation,the predictions of *experts* is weighed according to the corresponding *gates*.

The *experts* and *gates* of MCGSM take the form as below

$$
\begin{aligned}
p(y|x) &= \sum_{c,s} p(c,s|x)\, p(y|c,s,x) \\
p(c,s|x) &\propto |\lambda_{cs} K_c|^{\frac{1}{2}} exp\left(-\frac{1}{2}\lambda_{cs} x^T K_c x\right) \\
p(y|x,c,s) &= |M_c|^{\frac{1}{2}} exp\left(-\frac{1}{2}\lambda_{cs}(y - A_c x)^T M_c (y - A_c x)\right) / (2\pi)^{\frac{D}{2}}
\end{aligned}
$$

where $\lambda_{cs} > 0$ and M_c and K_c are positive definite and the above equation assumes the means of *GSM* are zero and weights of components are equal.

The more general form will be

$$
\begin{aligned}
p(c,s|x) &\propto \pi_{cs} |\lambda_{cs} K_c|^{\frac{1}{2}} exp\left(-\frac{1}{2}\lambda_{cs}(x - u_c)^T K_c (x - u_c)\right) \\
p(y|x,c,s) &= |M_c|^{\frac{1}{2}} exp\left(-\frac{1}{2}\lambda_{cs}(y - A_c x - m_c)^T M_c (y - A_c x - m_c)\right) / (2\pi)^{\frac{D}{2}}
\end{aligned}
$$

with $\pi_{cs} > 0$.This parametrization has quadratic growth in number of parameters with the dimension of x .So we perform low rank approximations.K_c is replaced with low rank approximaton

$$
K_c = \sum_n \beta_{cn}^2 b_n b_n^T
$$

using rank one basis matrices $b_n b_n^T$ which are shared over components also.The squaring of β preserves the positive definiteness of K_c.After reparametrizing,

$$
\begin{aligned}
p(c,s|x) &\propto exp\left(\eta_{cs} - \frac{1}{2} e^{\alpha_{cs}} \sum_n \beta_{cn}^2 \left(b_n^T x\right)^2 + e^{\alpha_{cs}} w_c^T x\right) \\
p(y|x,c,s) &= |L_c| exp\left(\frac{M}{2}\alpha_{cs} - \frac{1}{2} e^{\alpha_{cs}} (y - A_c x - m_c)^T L_c L_c^T (y - A_c x - m_c)\right) / (2\pi)^{\frac{M}{2}}
\end{aligned}
$$

The *gates* pick an expert based on the input x and *experts* are just *Gaussians* with linearly predicted mean.This completes the total formulation of MCGSM.Next we shall see denoising using MCGSM as an application.

2.2 Denoising using MCGSM

Consider denoising of a noisy image as an application of MCGSM,

$$y = x + n$$

y is the noisy image,x is the clean image and n is the noise.We now want to obtain estimate x from y.For this we use the *MAP* inference and *image prior* concept.We already had trained a statistical model (*image prior*) for clean images,and now given a noisy image,we want to steer it towards the *field* of clean images,but this is as ill-posed problem,we report that images which will deliver *maximum* aposterior $p(x|y)$ .This is the *MAP* inference method.We also use the *Gradient Ascent* to obtain the maximum of the *Log-likelihood.*

$$\begin{aligned} \hat{x} &= \underset{x}{argmax}\, p(x|y) \\ &= \underset{x}{argmax}\, p(x)\, p(y|x) \end{aligned}$$

$p(x)$ is the *prior* ,already modelled using the MCGSM and $p(y|x)$ can be calculated.It essentially is proprotional to $exp\left(- \parallel y - x \parallel^2 / \sigma^2\right)$ with σ being the noise standard deviation.

For maximising aposterior ,we consider maximising the *loglikelihood* as it simplifies the process.

$$\hat{x} = \underset{x}{argmax}\, log\,(p\,(x|y))$$

By incorporating the Gradient Ascent ,we get

$$\hat{x}_{t+1} = \hat{x}_t + \eta \nabla_x log\,(p\,(x)\,p\,(y|x))$$

$$\nabla_x log\,(p\,(y|x)) \propto \parallel y - x \parallel$$

and the derivative of the prior can be found out.This way we itaratively obtain cleaner images from the noise image.By setting appropriate learning rate,we can obtain cleaned images for different noise levels

2.3 Results

We denoised different images different noise levels and compared them with the *patch* based denoising techniques like *KSVD* .The results are as shown in the table

Figure	**noise level**	**MCGSM** psnr(dB)	**KSVD** psnr(dB)
barbara	15	30.91	31.51
	20	28.96	29.99
	25	27.5	28.81
	30	26.43	27.84
boat	15	31.10	31.96
	20	29.6	30.73
	25	28.21	29.79
	30	27.41	29.01
hill	15	30.80	31.71
	20	29.52	30.54
	25	28.25	29.68
	30	27.66	29.01
lena	15	33.01	34.03
	20	31.43	32.87
	25	30.4	31.95
	30	29.21	31.19

Table 1: Table comparing **MCGSM** with **KSVD**

(a) Original image (b) Denoised image

(c) Original image (d) Denoised image

Figure 2: Results

As we can observe in the results,MCGSM denoises well but the texture is not retained.We investigate reasons for this non-retainment of texture and try to overcome it in next section

Chapter 3 Adapting MCGSM

In this chapter we investigate and try ways for improving the MCGSM by adapting it to the specific image.

3.1 Modifying gradient

3.1.1 Gradient of Loglikelihood

As stated earlier we use *Gradient Ascent* technique to find the maximum of the *Log-likelihood* of the prior.

$$p(c,s|x) \propto exp\left(\eta_{cs} - \frac{1}{2}e^{\alpha_{cs}}\sum_{n}\beta_{cn}^{2}\left(b_{n}^{T}x\right)^{2} + e^{\alpha_{cs}}w_{c}^{T}x\right)$$

We consider $f(c,s|x) = exp\left(\eta_{cs} - \frac{1}{2}e^{\alpha_{cs}}\sum_{n}\beta_{cn}^{2}\left(b_{n}^{T}x\right)^{2} + e^{\alpha_{cs}}w_{c}^{T}x\right)$ Now to obtain a probability distribution $p(c,s|x)$we need to normalize $f(c,s|x)$with respect to c,s .

$$p(c,s|x) = \frac{f(c,s|x)}{\sum_{c,s} f(c,s|x)}$$

Now the *prior distribution* becomes

$$p(y|x) = \frac{\sum_{c,s} f(c,s|x)\, p(y|c,s,x)}{\sum_{c,s} f(c,s|x)}$$

Consequently *log-likelihood* (L)*will be*

$$\begin{aligned} L &= log\, p(y|x) \\ &= log\sum_{c,s} f(c,s|x)\, p(y|c,s,x) - log\sum_{c,s} f(c,s|x) \end{aligned}$$

Lets now take *derivative* of L wrt x and y .

$$\begin{aligned}
\frac{\partial L}{\partial x} &= \frac{\sum_{c,s} f\left(c,s|x\right)p\left(y|c,s,x\right)\left\{\left(-e^{\alpha_{cs}}\sum_{n}\beta_{cn}^{2}\left(b_n^T x\right)b_n + e^{\alpha_{cs}}w_c\right) + \left(e^{\alpha_{cs}}A_c L_c L_c^T\left(y - m_c - A_c^T x\right)\right)\right\}}{\sum_{c,s} f\left(c,s|x\right)p\left(y|c,s,x\right)} \\
&+ \frac{\sum_{c,s} f\left(c,s|x\right)\left\{e^{\alpha_{cs}}\sum_{n}\beta_{cn}^{2}\left(b_n^T x\right)b_n + e^{\alpha_{cs}}w_c\right\}}{\sum_{c,s} f\left(c,s|x\right)} \\
\frac{\partial L}{\partial y} &= -\frac{\sum_{c,s} f\left(c,s|x\right)p\left(y|c,s,x\right)\left\{e^{\alpha_{cs}}LL_c^T\left(y - m_c - A_c^T x\right)\right\}}{\sum_{c,s} f\left(c,s|x\right)p\left(y|c,s,x\right)}
\end{aligned}$$

3.1.2 Modification

The modification done is as follows,we pass the *clean image input* in the gate part of the *prior* intead of noisy image and leave the *expert* unmodified.

$$p\left(y|x\right) = \sum_{c,s} p\left(c,s|x_{clean}\right)p\left(y|c,s,x\right)$$

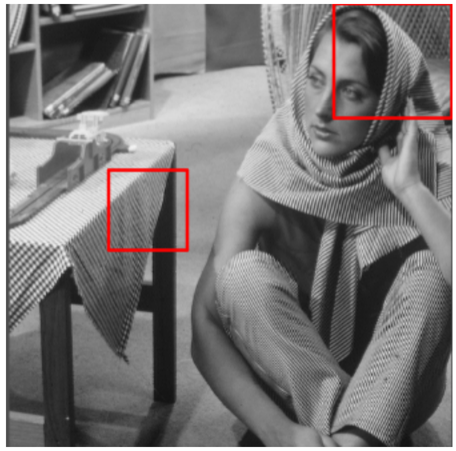
Consequently the *gradient* gets modified as below, as the *expert* still has x_{noisy} as the input so the derivative of it will be zero,

$$
\begin{aligned}
\frac{\partial L}{\partial x} &= \frac{\sum\limits_{c,s} f\left(c,s|x_{clean}\right) p\left(y|c,s,x\right)\left\{\left(-e^{\alpha_{cs}}\sum\limits_{n}\beta_{cn}^{2}\left(b_n^T x_{clean}\right) b_n + e^{\alpha_{cs}} w_c\right)\right\}}{\sum\limits_{c,s} f\left(c,s|x_{clean}\right) p\left(y|c,s,x\right)} \\
&+ \frac{\sum\limits_{c,s} f\left(c,s|x_{clean}\right)\left\{e^{\alpha_{cs}}\sum\limits_{n}\beta_{cn}^{2}\left(b_n^T x_{clean}\right) b_n + e^{\alpha_{cs}} w_c\right\}}{\sum\limits_{c,s} f\left(c,s|x_{clean}\right)} \\
\frac{\partial L}{\partial y} &= \frac{\sum\limits_{c,s} f\left(c,s|x_{clean}\right) p\left(y|c,s,x\right)\left\{e^{\alpha_{cs}} L L_c^T\left(y - m_c - A_c^T x_{clean}\right)\right\}}{\sum\limits_{c,s} f\left(c,s|x_{clean}\right) p\left(y|c,s,x\right)}
\end{aligned}
$$

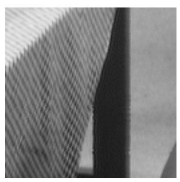
The motivation behind doing such modification is to see and analyze if the selection of component is the problem and whether noise is preventing it.Since selection of appropriate component is very crucial for retaining the texture,we passed the clean image input to prevent noise from making a wrong selection if at all

3.1.3 Results

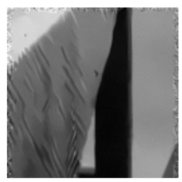
We performed image denoising similar to as done by using MCGSM but here we use the modified gradient instead.The results are shown below



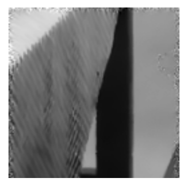
(a) Original Image



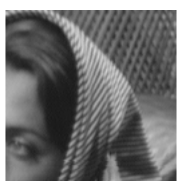
(b) Original patch



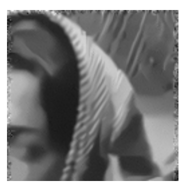
(c) Denoised using model with noisy input psnr=26.8



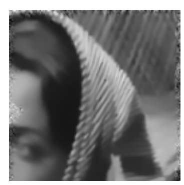
(d) Denoised using model with clean input psnr=27.55



(e) Original patch



(f) Denoised using model with noisy input psnr=27.3



(g) Denoised using model with clean input psnr=28.26

Figure 3: results

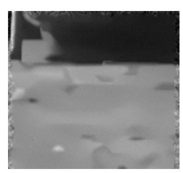
(a) Original Image

(b) Original patch

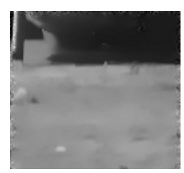
(c) Denoised using model with noisy input psnr=27.88

(d) Denoised using model with clean input psnr=28.12

(e) Original patch



(f) Denoised using model with noisy input psnr=28.5



(g) Denoised using model with clean input psnr=28.97

Figure 4: results

As we can see, visual quality of the images denoised with modified gradient is only slightly improved when compared to the original gradient.This states that,the selection of component is not the problem,as we passed the ground truth image itself.The entire component which would explain the existing texture is missing.

This is reasonable,as the generic model is trained on a different dataset,it may not contain the exact component which would explain the specific texture in the image considered.We need to learn new components specific to the given image.

3.2 Modifying parameters

3.2.1 Theory

As we saw in the previous section,the actual problem is absence of appropriate components in the generic model.So in this section we try to modify and adapt some paramters of the generic model on the given image.We use clean image itself for modifying as an analysis.

In the model of MCGSM ,$p(y|x) = \sum_{c,s} p(c,s|x)\, p(y|c,s,x)$ the expert $p(y|c,s,x)$is a *Normal* Distribution in y with linearly predicted mean ,dependent on causal neighbourhood x

$$p(y|x,c,s) = K\, exp\left(-\frac{1}{2}\lambda_{cs}(y - A_c x)^T M_c (y - A_c x)\right)$$

So now we retrain the whole model by only modifying *predictors* A_c as they are the ones that are more image specific than others.Other parameters are kept as they are.

3.2.2 Results

Images are denoised using the retrained model and they are compared with the images denoised using the generic model.The results are shown below

(a) denoised image using generic model psnr=28.21

(b) Denoised image using retrained model psnr=28.49

(c) denoised image using generic model psnr=28.25

(d) Denoised image using retrained model psnr=28.31

Figure 5: Results

The images denoised using the retrained model are very slightly better than the denoised images using generic model.They still have the problem of texture smoothening unsolved.As the clean image itslef is used for retraining of the model,the reason for very slight improvement is,as the generic model is trained on a large set of various images,the parameters of the generic model are stable and a small retraining of the parameters using the specific image is not strong enough to adapt the parameters of the generic model.

So we need to train a whole new set of components for the specific image.We shall see this in the next section.

3.3 Training new model

3.3.1 Theory

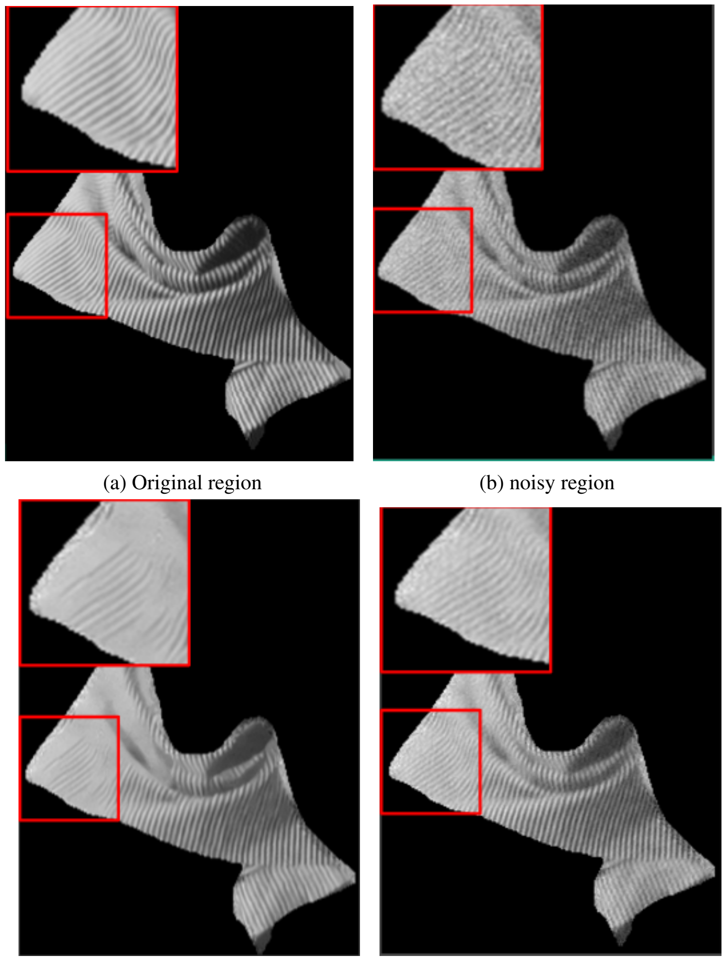
As said in the previous section,we need to train a whole new set of components and scales for effective explanation of the textures in image.For this we consider similar clusters of the image and train a model on each of them and denoise the region using that particular model.Whole image is divided into such similar regions

The reason for collecting similar regions is that,each component is said to explaining a different feature,like horizontal edges,slant edges and so on.So by selecting similar regions,we collect regions having similar such features,which makes the components easy for explaining those features.

Denoising is done separately on each of those similar regions using the model trained on that region.

3.3.2 Results

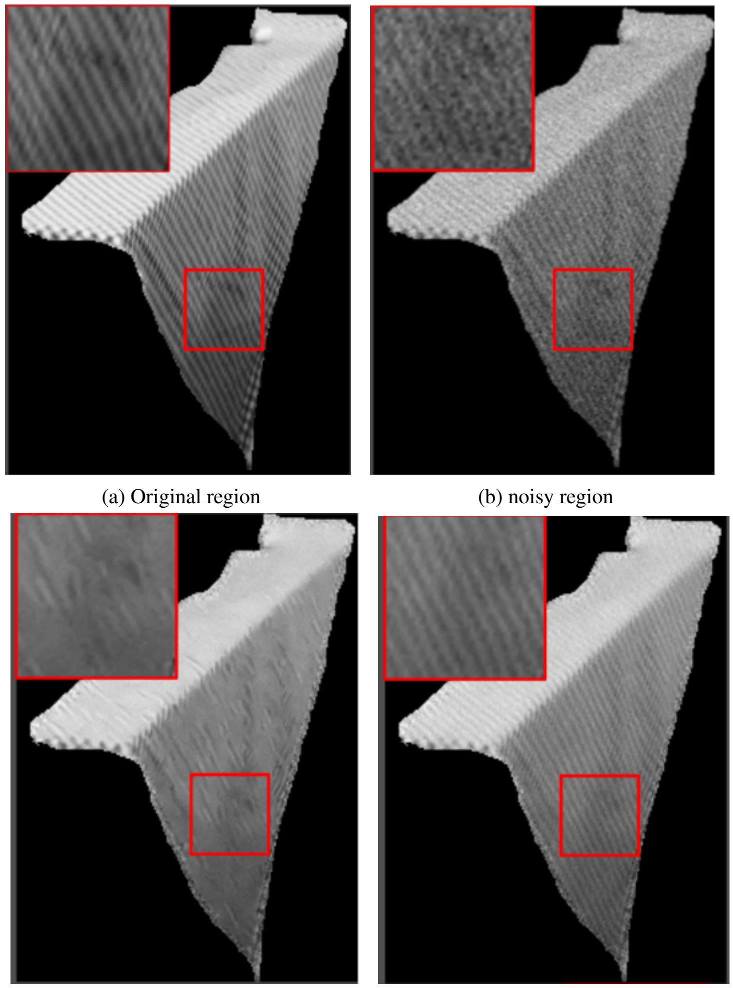
We collect similar regions as said earlier and train a model and denoise them using that model.Denoise results are shown below



(a) Original region (b) noisy region

(c) Denoised using generic model (d) Denoised using trained model

Figure 6: results



(a) Original region (b) noisy region

(c) Denoised using generic model (d) Denoised using trained model

Figure 7: results

As we can see,clearly the images denoised using the model trained on noisy data are much better than images denoised using generic model.So this says that extractinf similar regions from the image and training a model on each one of them has learnt the components which would explain the texture well.This learning of textures has led to preserving the texture of images while denoising which is a serious drawback of MCGSM.Thus we were able to overcome the texture smothening problem of MCGSM.