

Non-Linear Restoration from a Single Frame Super Resolution Using Pearson Type VII Density

Sakinah Ali Pitchay *

Abstract—Super-resolution seeks to recover a high resolution image from one or more low resolution images. It is an ill-posed problem, with no consensus how best to devise image models that can both improve smoothness and preserve the edges in the image. Here we investigate the use of prior based on Pearson type VII density integrated with a Markov Random Field (MRF) model. We formulate two different versions, one that acts on the pixel level and another one that acts on the entire image. Having a single down-sampled and noisy version of low resolution frame, we aim to obtain the high resolution image. We compare the state of the art of image priors in super resolution application and we discover that our image prior Pearson-MRF achieves the best performance in terms of qualitative measurement.

Keywords: single-frame super-resolution, Pearson type VII, MRF model

1 Introduction

Super-resolution seeks to generate a high resolution image from one or more low resolution images. The limitations of the capturing source often allow the loss of resolution including the shifting, rotation, blur and down-sampling. Moreover, the capturing process instigates additive noise that causes it is not sufficiently to sample the scene adequately. Often the observed frames are deficient or noisy, which makes this problem ill-posed and possibly under-determined too. Thus, extra knowledge is vital to acquire an adequate solution and well-known as image prior. Employing the probabilistic model based framework, this extra information may be specified as a prior distribution on the salient statistics that images are known to have. The two main criterions are apparently contrary each other: local smoothness and the existence of edges. Hence the requirement of a good image prior is demanding.

The former prior models have been proposed in the lit-

*Sakinah Ali Pitchay is with the Faculty of Science and Technology, Islamic Science University of Malaysia (USIM), Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia. Currently she is a PhD student in School of Computer Science, The University of Birmingham, United Kingdom. (email:S.A.Pitchay@cs.bham.ac.uk / sakinah.ali@usim.edu.my).

erature, yet with no substantiation. Gaussian Markov Random Fields represent a common choice for its computational tractability. The Huber-MRF is prominent since it is more robust but still convex and works in [5, 6, 7] are considered to be the state of the art approach.

In our previous paper [3], we proposed a robust density, the univariate version of Pearson type VII formulated as Markov Random Field(MRF) in super resolution approach. Previously, the comparison with the existing image priors are concentrating on compressive matrices transformation. Due of curiosity, we formulated and examined the multivariate of Pearson type VII and compare it with the state of the art approach using the classical super resolution technique. This density is formerly used as robust density estimation in [1] as alternative to the t-mixtures and in stock market modelling [2]. The remainder of the paper is organised as follow. In Section 2, we describe the problem formally including how to estimate the high resolution image. Section 3 presents the image prior that we investigate. Section 4 presents our proposed solution and Section 5 details the experimental results and their analysis. Finally, Section 6, concludes the paper.

2 Framework of Image Super Resolution

2.1 Observation Model

The high resolution image of $N = r \times c$ pixel intensities will be vectorised and denoted as \mathbf{z} . This image suffers a quite complicated transformation into a low-resolution frame includes blur and down-sampled. We adopt a linear model to express this transformation which, although it is not completely accurate, it has worked well in many previous studies on super-resolution [4, 5, 7]. Denoting the low resolution frame by \mathbf{y} in a vectorised form, and the linear transform that takes \mathbf{z} into \mathbf{y} by \mathbf{W} where it is a stack of transformation matrices into a single matrix \mathbf{W} . We expand the forward model as the following:

$$\mathbf{y} = \mathbf{W}\mathbf{z} + \eta \quad (1)$$

\mathbf{W} is a product of blurring and down-sampling matrix of size $[\mathbf{M} \times \mathbf{N}]$, usually ill-conditioned matrix that models

a linear blur operation and the down-sampling by row and column operator. $\boldsymbol{\eta}$ is a vector that represents an additive noise, assumed to be Gaussian with zero-mean and σ^2 variance. To make the problem more challenging, an additive noise is contaminated to the blurred and down-sampled image.

2.2 The Joint Model

The overall model is the joint model of the observations \mathbf{y} and the unknowns \mathbf{z} . Using these, we have joint probability

$$Pr(\mathbf{y}, \mathbf{z}) = Pr(\mathbf{y}|\mathbf{z})Pr(\mathbf{z}) \quad (2)$$

where the first term is the observation model, the second term is the image prior model. Hence we have for the first term in (2):

$$Pr(\mathbf{y}|\mathbf{z}) \propto \exp \left\{ -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{W}\mathbf{z})^T(\mathbf{y} - \mathbf{W}\mathbf{z}) \right\} \quad (3)$$

This is also called the model *likelihood*, because it expresses how likely it is that a given \mathbf{z} produced the observed \mathbf{y} through the transformation \mathbf{W} . The second term of (2) will be instantiated with either one of the image priors discussed in Section 3. To achieve our goal, we need to 'invert' the causality described by our model, to infer the latent variables \mathbf{z} from the observed variables \mathbf{y} .

2.3 Inverting the model to estimate \mathbf{z}

We can invert the causality encoded in a probabilistic model by the use of Bayes' rule.

$$Pr(\mathbf{z}|\mathbf{y}) = \frac{Pr(\mathbf{y}|\mathbf{z})Pr(\mathbf{z})}{Pr(\mathbf{y})} \quad (4)$$

This is called the *posterior* probability of \mathbf{z} given the observed data \mathbf{y} . Eq. (4) says that, the probability that \mathbf{z} is the hidden image that gave rise to what we observed, i.e. \mathbf{y} , is proportional to the likelihood that this \mathbf{z} fits the data \mathbf{y} and the probability that this bunch of N intensity values, i.e. the vector \mathbf{z} , actually 'looks like' a valid image. Note that the latter is desperately needed in underdetermined systems, since there are infinitely many vectors \mathbf{z} that fit the data.

2.4 Maximum A Posteriori Inference through Optimisation

To obtain the most probable estimate of \mathbf{z} that conforms to our model and data, we need to maximise (4) as a function of \mathbf{z} . Observe that, the denominator, $Pr(\mathbf{y})$ does not depend on \mathbf{z} . Hence, the maximum value of the fraction (4) occurs for exactly the same \mathbf{z} for which the maximum of the numerator does. That is, the most probable estimate is given by:

$$\hat{\mathbf{z}} = \arg \max_{\mathbf{z}} \frac{Pr(\mathbf{y}|\mathbf{z})Pr(\mathbf{z})}{Pr(\mathbf{y})} \quad (5)$$

$$= \arg \max_{\mathbf{z}} Pr(\mathbf{y}|\mathbf{z})Pr(\mathbf{z}) \quad (6)$$

Further, this maximisation is also equivalent to maximising the logarithm in the right hand side, since the logarithm is a monotonic increasing function. We can also turn the maximisation into minimisation, by flipping the signs, as in the following equivalent rewriting:

$$\hat{\mathbf{z}} = \arg \min_{\mathbf{z}} \{-\log[Pr(\mathbf{y}|\mathbf{z})] - \log[Pr(\mathbf{z})]\} \quad (7)$$

In words, the most probable high resolution image is the one for which the negative log of the joint probability model takes its minimum value. Thus, our problem is now solvable by performing this minimisation. The expression to be minimised, i.e. the negative log of the joint probability model may be interpreted as an error objective, and shall be denoted as:

$$Obj(\mathbf{z}) = -\log[Pr(\mathbf{y}|\mathbf{z})] - \log[Pr(\mathbf{z})] \quad (8)$$

The most probable estimate is the $\hat{\mathbf{z}}$ that has highest probability in the model. Equivalently the one that achieves the lowest error. Since our model has had two factors (the likelihood or observation model, and the image prior), consequently our error-objective also has two terms: the misfit to observed data, and 'penalty' for violating the smoothness and/or other characteristics encoded in the prior. By plugging in the functional forms for the observation model and for the various possible priors into (8), we now give the specific form of this objective function below, so the interpretation of the individual error terms is more evident. We will make use of the following notation, taking the log of eq. (3):

$$l(\mathbf{z}) := -\log Pr(\mathbf{y}|\mathbf{z}) + const. \quad (9)$$

$$= \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{W}\mathbf{z})^T(\mathbf{y} - \mathbf{W}\mathbf{z}) \quad (10)$$

3 Prior Image Model: Markov Random Fields

The main characteristic of any natural image is a local smoothness. That is, the intensities of neighbouring pixels tend to be very similar. A MRF is a joint distribution over all the pixels on the image that captures spatial dependencies of pixel intensities. A first-order MRF assumes that, for any pixel, its intensity depends on the intensities of its closest cardinal neighbours but does not depend on any other pixel of the image. Here we will adopt the 1-st order MRF that conditions each pixel of intensity on its four cardinal neighbours in the following way. For any one pixel z_i we define:

$$Pr(z_i|\mathbf{z}_{-i}) = Pr(z_i|z_{4\text{neighb}(i)}) \quad (11)$$

$$= Pr(z_i - \frac{1}{4} \sum_{j \in 4\text{neighb}(i)} z_j) \quad (12)$$

where the notation \mathbf{z}_{-i} means all the pixels excluding the i -th, and the set of four cardinal neighbours of z_i was

denoted as $4\text{neighb}(i)$. This is a univariate probability distribution.

Consequently, for the whole image of N pixels, the MRF represents the joint probability over all the pixels on the image — a multivariate probability distribution.

$$Pr(\mathbf{z}) \propto \prod_{i=1}^N Pr(z_i | z_{4\text{neighb}(i)}) \quad (13)$$

$$= \prod_{i=1}^N Pr\left(z_i - \frac{1}{4} \sum_{j \in 4\text{neighb}(i)} z_j\right) \quad (14)$$

The notation '∝' means 'proportional to', i.e. there is a division by a constant that makes the probability density integrate to one. This constant may depend on various parameters of the actual instantiation of the building block probability densities, but it does not depend on \mathbf{z} . Since in this work we only need to estimate \mathbf{z} , therefore we can ignore the expression of the normalising constant throughout.

This form of MRF has been previously employed with success in e.g. [4, 5]. Alternatives include the so-called total variation model, employed e.g. in e.g. [7], which is based on image gradients, also quite simple. In [6], an experimental comparison of these two alternatives suggests these have comparable performance, the former being slightly superior though.

The simplicity of (14) is also intuitively appealing. One can think of the difference between a pixel intensity and the average intensity of its neighbours, i.e. $z_i - \frac{1}{4} \sum_{j \in 4\text{neighb}(i)} z_j$, as a *feature*. Considering that we want to encode the general smoothness property of images, it is easy to see that this feature is very useful: Whenever this difference is small in absolute value, we have a smooth neighbourhood. Whenever it is large in absolute value, we have a discontinuity. Hence, to express smoothness, we just need to instantiate the probability distribution over this feature, i.e. the uni-variate densities in the product (14), $Pr(z_i - \frac{1}{4} \sum_{j \in 4\text{neighb}(i)} z_j)$, with symmetric densities around zero, which give high probability to small values. The Gaussian is a good example. In the same time, to allow for a few discontinuities, we need to use heavy tail densities, such as the Huber or the Pearson type VII density.

To simplify notation and it is conveniently to create the symmetric $N \times N$ matrix \mathbf{D} to encode the above neighbourhood structure, with entries:

$$d_{ij} = \begin{cases} 1 & \text{if } i = j; \\ -1/4 & \text{if } i \text{ and } j \text{ are neighbours;} \\ 0 & \text{otherwise.} \end{cases}$$

Then we may write the i -th feature in a vector form, with the aid of the i -th row of this matrix (denoted \mathbf{D}_i) as the

following:

$$z_i - \frac{1}{4} \sum_{j \in 4\text{neighb}(i)} z_j = \sum_{j=1}^N d_{ij} z_j \quad (15)$$

$$= \mathbf{D}_i \mathbf{z} \quad (16)$$

Again, this is the i -th neighbourhood feature of the image, and there are $i = 1, \dots, N$ such features on an N -pixel image.

The studies of data visualisation of the neighbourhood features ($\mathbf{D}_i \mathbf{z}$) from several natural images are presented in a histogram. We now turn to instantiate the functional form of the probability densities that describe the shape of the likely values of these features. Figure 1 shows a few examples of observed histograms of these features, from natural images. The probability densities that we employ in our image priors should ideally have similar shapes.

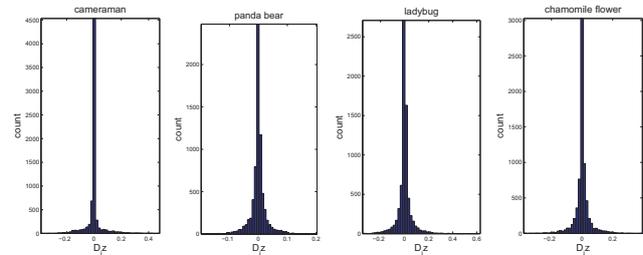


Figure 1: Examples of histograms of the distribution of neighbourhood features $\mathbf{D}_i \mathbf{z}$, $i = 1, \dots, N$ from natural images.

3.1 Gaussian-MRF

The Gaussian MRF is the most widely used image prior density. It has the following form:

$$Pr(\mathbf{z}) \propto \prod_{i=1}^N \exp\left\{-\frac{1}{2\lambda} (\mathbf{D}_i \mathbf{z})^2\right\} \quad (17)$$

$$= \exp\left\{-\frac{1}{2\lambda} \sum_{i=1}^N (\mathbf{D}_i \mathbf{z})^2\right\} \quad (18)$$

where λ is the variance parameter.

3.2 Huber-MRF

The Huber density is defined with the aid of the Huber function. It takes a threshold parameter δ , specifying the value at which it diverts from being quadratic to being linear. A generic variable u in the definition of this function will be instantiated later as a neighbourhood-feature $\mathbf{D}_i \mathbf{z}$ within the image prior use.

$$H(u|\delta) = \begin{cases} u^2, & \text{if } |u| < \delta \\ 2\delta|u| - \delta^2, & \text{otherwise.} \end{cases} \quad (19)$$

The Huber-MRF prior is then defined in (21) where λ is similar to a variance parameter.

$$Pr(\mathbf{z}) \propto \prod_{i=1}^N \exp \left\{ -\frac{1}{2\lambda} H(\mathbf{D}_i \mathbf{z} | \delta) \right\} \quad (20)$$

$$= \exp \left\{ -\frac{1}{2\lambda} \sum_{i=1}^N H(\mathbf{D}_i \mathbf{z} | \delta) \right\} \quad (21)$$

4 Pearson Type VII-MRF

4.1 The univariate Pearson Type VII-MRF

The Pearson-MRF made of univariate building blocks: A zero mean univariate Pearson prior, is defined as:

$$Pr(\mathbf{z}) \propto \prod_{i=1}^N \{ (\mathbf{D}_i \mathbf{z})^2 + \lambda \}^{-\left(\frac{1+\nu}{2}\right)} \quad (22)$$

where ν and λ control the shape of the distribution.

4.2 The multivariate Pearson Type VII-MRF

A zero mean multivariate Pearson-MRF density in a generic N-dimensional vector of $\mathbf{D}_i \mathbf{z}$, has the following form:

$$Pr(\mathbf{z}) \propto \left\{ \sum_{i=1}^N (\mathbf{D}_i \mathbf{z})^2 + \lambda \right\}^{-\left(\frac{\nu+N}{2}\right)} \quad (23)$$

4.3 Discussion on the two versions of Pearson-MRF

The version devised in Section 4.1 may be regarded as having independent Pearson-priors on each neighbourhood-feature. Of course, we ought to point out that the neighbourhood features are not independent in reality. However, since each pixel only depends on four others, it may be a reasonable approximation.

The version gave in section 4.2, in turn, does not allow such independence interpretation. Conversely, this can has the advantage that the spatial dependencies are not broken up, but more reliably accounted for. However, on the downside, the heavy tail behaviour is more advantageous to have on the pixel level, i.e., on the distribution of neighbourhood features. Indeed, it is the distribution of neighbourhood features the one in which the edges from the image creates outliers. In turn, the multivariate Pearson-MRF is a density on images. Hence, its heavy tail behaviour would be well suited to account for outlying or atypical images. Including both of these versions in our comparison will therefore uncover to us which of these pros or cons are more important for recovering quality high resolution images.

5 Experiments

5.1 Experimental Setting

We present two set of a single frame image super resolution experiments illustrating the performance of the hyper-parameters and four different images for testing the stability of the parameters found. The LR image is blurred by the uniform blur matrix of size 3x3, down-sampled by factor 4 and contaminated by standard deviation of Gaussian noise of 1e-3 and 1e-2. The original image for Exp. 1 is the cameraman, Exp. 2 is the panda bear, Exp. 3 is the chamomile flower and Exp 4 is the ladybug. All are in size 100x100. The initial guess was initialized with Gaussian-MRF with σ^2/λ set to 1 and was used as a starting point for the recovery algorithm. We employed a conjugate gradient type method¹, which requires the gradient vector of the objectives. All the pixel intensities are scaled to interval [-0.5,0.5].

In our paper [3], we applied the compressive matrix of W to find out how well is the proposed image prior based MRF in comparison with other state of the art priors, and hyper-parameters is manually tuned to acquire the optimum mean square error for all methods. Meanwhile in this paper we compare it using the transformation matrix consists of blur and down-sampling, which one of the two version of Pearson works better and does it still good although for a different transformation? Secondly how the selection of hyper-parameters of univariate Pearson type VII is determined and how good is the proposed prior in under-determined system where W has [2500,10000] size. For this experiment, we wider the test images and the noise variance as well.

5.2 Hyper-parameters Selection of Pearson-MRF

The performance of the image recovery of high resolution is depending on how good selection value of hyper-parameters in image prior. Bad estimation can lead to produce a bad result. Since we are assessing the performance of both parameters, the recovery algorithm is assuming knowing the true noise variance σ^2 . From the observation using the constructed blur and down-sampling matrix W, we found practical range of λ and ν .

The results are presented in Figure 2. Too small λ (0.001) and ν reduces the effect of prior and the solution approaching the Maximum Likelihood(ML). Whilst too big λ (10000) will blur the edges. The overall performance of the recovered image is depending strongly on the selection of λ . We can conclude that the ν can be fixed into a practicable range (i.e:1-10) so that the iteration could terminate earlier and the λ is found best from 0.1 to 100. Two set of images (cameraman and panda image) are

¹We made use of the efficient implementation available from <http://www.kyb.tuebingen.mpg.de/bs/people/carl/code/minimize/>

examined to achieve the best performance. While Figure 3 shows the variation performance varying several λ . Besides, the performance for several level of noise is investigated using one of the stable range of ν using four different images and the results are presented in Figure 4.

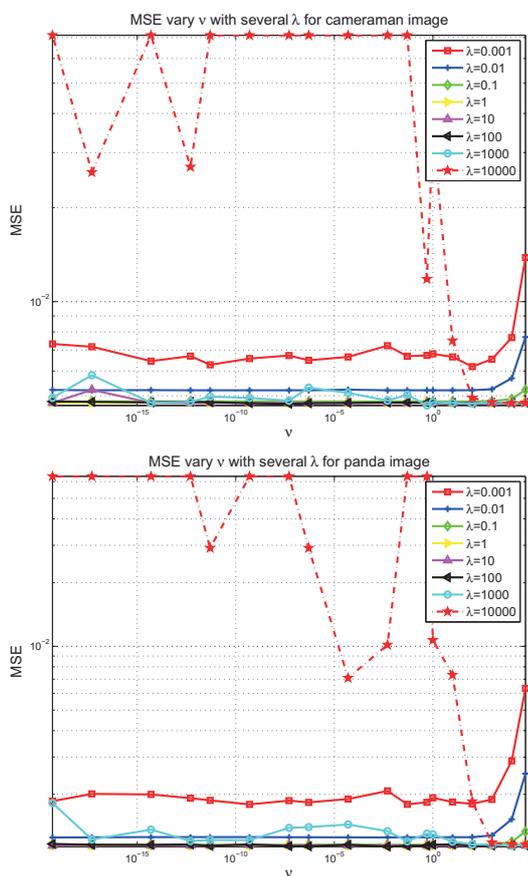


Figure 2: Top: Test image of cameraman, bottom: test image of panda are used to inspect the best value of hyper-parameters by computing the MSE performance varying several λ where noise variance is 0.001.

5.3 Results

To assess the goodness of the proposed method, Pearson-MRF estimation results are compared with image enhancement state of the art methods in [4, 5, 7] using the qualitative measurement, mean square error (MSE). The competing image priors are: Gaussian-MRF, a multivariate Pearson type VII based MRF and the Huber-MRF. These results are presented in Figure 5 and we can see that the univariate Pearson type VII based MRF can achieve state-of-the-art performance, comparable to the Huber-MRF as well across all noise levels tested while the other priors tested perform worse. Finally we also illustrated two recovery set of experiment where the transformation matrix consists blur and down-sampling. Here we generated a single frame of high resolution for under-

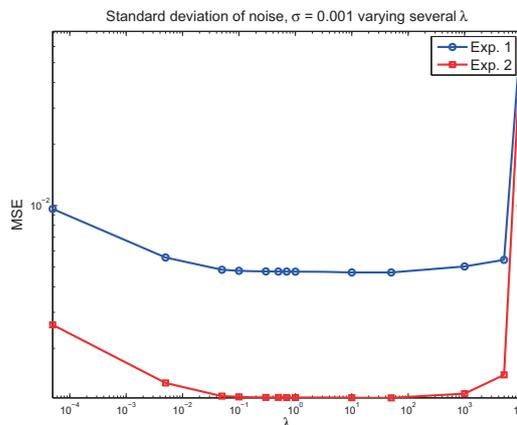


Figure 3: MSE measurement varying λ where the ν is fixed to 0.05 and this value is found one of the best from manually search.

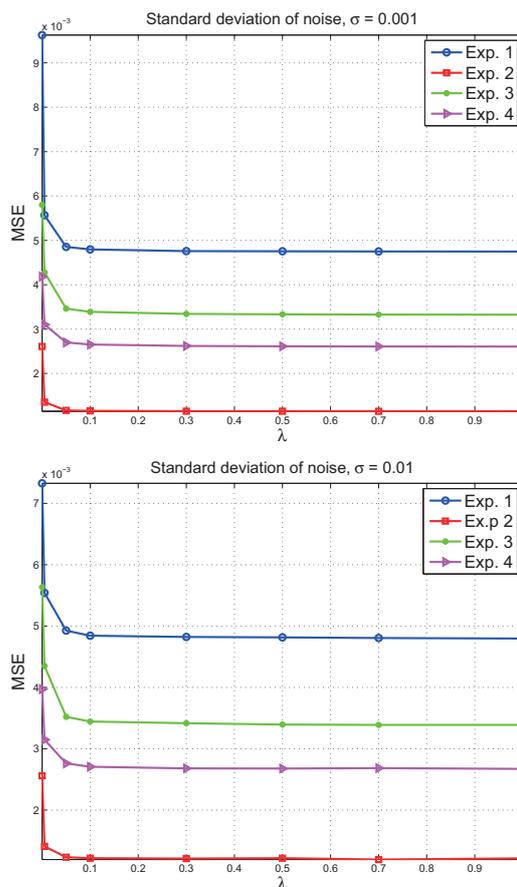


Figure 4: Test set on a different level of noise for four type of images varying several λ using one of the optimal value found ($\nu = 0.05$).

determined system as well in this case. Figure 7 shows the outcome using univariate Pearson type VII based MRF.

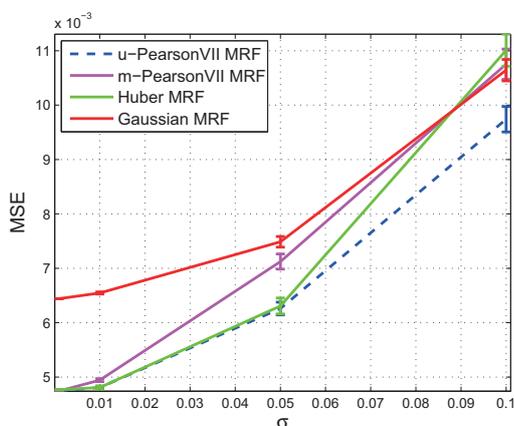


Figure 5: Comparative MSE performance for under-determined system where W has $[2500,10000]$ and varying several level of noise using the best values of hyperparameter for every image prior. The error bars are over 10 independent trials.

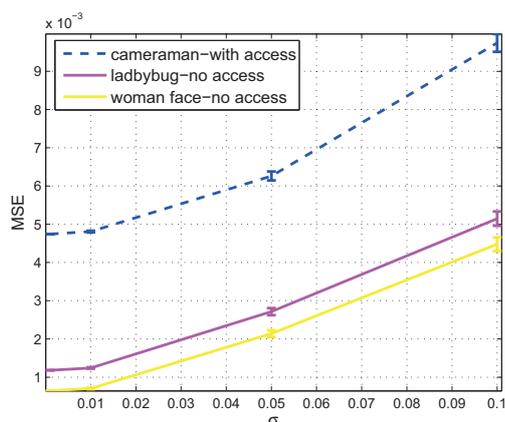


Figure 6: Comparing the MSE performance of the image with the access to the ground truth image for finding the hyperparameters Pearson type VII baased MRF and using the same value of the optimal found for other images. The error bars are over 10 independent trials over the random draw of the additive noise and the transformation W consists of blur and down-sampling.

6 Conclusions and Future Work

In this paper we formulated two versions of Pearson-MRF image priors, and conducted a comparative experimental study among these and state of the art methods of image prior from a single noisy version of low resolution image. We demonstrate that our proposed prior, univariate Pearson Type VII-MRF is likewise superior with Huber for all level of noise. The recovered image is always consistent although it has several local optima and we asses

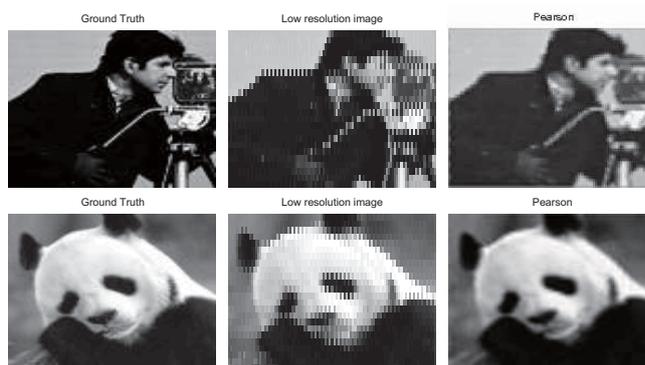


Figure 7: Left: Ground truth image, blurred and down-sampled image corrupted by additive noise and estimated image using Pearson-MRF from a noisy version of single low-resolution frame. The problem is under-determined system where $W[2500,10000]$ and the σ is 0.001.

four different images. Our motivation for Pearson-MRF prior has been the heavy tail property of the Pearson type VII-distribution, which indeed seems to be a good way of preserving the edges too while imposing smoothness. We tested this in under-determined systems, using the optimal value under various natural images. Future work is aimed towards recovering from multiple frames and working with multiple scenes for under-determined system and over-determined system as well.

References

- [1] J.Sun, A.Kabán and J.Garibaldi, "Robust Mixture Modeling using the Pearson Type VII Distribution", *Proc. Int. Joint Conference on Neural Network*, 2010, to appear. <http://www.cs.bham.ac.uk/~axk/PearsonTypeVIIMixture.pdf>
- [2] Y.Nagahara, "Non-gaussian distribution for stock returns and related stochastic differential equation", *Asia-Pacific Financial Markets*, V3, N2, pp. 121-149, 1996
- [3] AKaban and S.A. Pitchay, "Single-frame Image Super-resolution Using a Pearson Type VII MRF", *Proc. IEEE International Workshop on Machine Learning for Signal Processing*, (MLSP 2010)
- [4] R. C. Hardie, K. J. Barnard, "Joint MAP Registration and High-Resolution Image Estimation Using a Sequence of Undersampled Images", *IEEE Trans. Image Processing*, V6, N12, Dec. 1997, pp. 621-633.
- [5] H. He and L.P. Kondi, "MAP Based Resolution Enhancement of Video Sequences Using a Huber-Markov Random Field Image Prior Model", *IEEE Conference of Image Processing*, 2003, pp. 933-936.
- [6] H. He and L.P. Kondi, "Choice of Threshold of the Huber-Markov Prior in MAP Based Video Resolution Enhancement", *IEEE Electrical and Computer Engineering Canadian Conference*, 2004.
- [7] L. C. Pickup, D. P. Capel, S. J. Roberts, A. Zissermann, "Bayesian Methods for Image Super-Resolution", *The Computer Journal*, 2007.