

Single Frame Image Recovery for Super Resolution with Parameter Estimation of Pearson Type VII Density

Sakinah Ali Pitchay

Abstract—Image recovery of super resolution aims to recover a single high resolution image from one or more low resolution frames. It is an ill-posed problem when the solution does not exist or it is unique. Thus, we introduce the prior based on Pearson type VII density integrated with a Markov Random Field (MRF) model. We devise two different versions, one that acts on the pixel level and another one that acts on the entire image. Here we present our parameter estimation and evaluate our approach using qualitative measurement in both compressive measurement and classical super resolution. Our estimation is theoretically simple and easy to implement.

Index Terms—single frame super resolution, Pearson type VII, MRF model, compressive measurement

I. INTRODUCTION

IMAGE recovery 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 known as image prior.

Using 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 one another: local smoothness and the existence of edges. To solve this particular problem, an investigation on a density function that has the ability to recover the image which allows for greater variability by having larger tails than the standard normal distribution (i.e:Gaussian). This density must be robust or has the heavy tail property so that it would be able to cope with the outlier. Hence the requirement of a good image prior is demanding.

The former prior models have been proposed in the literature, yet with no substantiation. Gaussian Markov

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, University of Birmingham, United Kingdom. email:S.A.Pitchay@cs.bham.ac.uk / sakinah.ali@usim.edu.my

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 [6], [7], [8] are considered to be the state of the art approach. They employ Huber prior, however the threshold is fixed in [6], [8].

In our recent work [3], we study and compare the state of the art of image priors in conventional super resolution application using manual selection from several search space. We extend our work by estimating the hyper-parameters of Pearson prior using held out estimation and cross validation. This includes other image priors as well. Besides that, we test our estimation parameter in both transformation (i) compressive measurement and (ii) classical super resolution. Indeed, the ground truth image is not accessible and the successful works in random property [10] inspired us to cope with it. We exploit the compressive measurement using real image.

Previously, we proposed a robust density, the univariate version of Pearson type VII formulated as Markov Random Field (MRF) in super resolution approach [4]. The comparisons 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 II, we describe the problem formally including how to estimate the high resolution image. Section III presents the image prior that we investigate for this experiment. The Pearson type VII based MRF is described in Section IV. Example results on automated estimation with comparisons to other image priors are presents in Section V. It depicts the experimental setting and its discussion. Finally, conclusions and future work are discussed in Section VI.

II. FRAMEWORK OF IMAGE RECOVERY SUPER RESOLUTION

In order to solve super-resolution problem, we must formalise it first. We will employ the probabilistic

formalism, which is well suited for its principled nature and its flexibility. Firstly, we need to construct a model of how the (unknown) high resolution image might have given rise to the (observed) low resolution images. This is often referred to as the *forward model* or *observation model*. This model serves as a formal abstraction of the real physical process: the high-resolution image is the 'cause', the low resolution images are the 'effects'. Since we observe the 'effects' but want to find out the 'cause', we then need to invert this model. This is a *backward operation*, often termed as *inference*.

As already mentioned, this is easier said than done for several reasons: (a) we may need to infer more pixel intensities than we have observed ones in the first place; (b) noise on the low resolution images degrades their information content. Hence, the forward model and the data alone is always insufficient. Fortunately, there is a prior knowledge about the statistics of natural images that we can exploit. Therefore the second part of the overall model for super-resolution is a model of a (generic) high-resolution image, encoded as a *prior*.

A. Model Formulation

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 [5], [6], [8]. Denoting the k -th low resolution frame by \mathbf{y}_k in a vectorised form, and the linear transform that takes \mathbf{z} into \mathbf{y}_k by \mathbf{W}_k , we can write the forward model as the following:

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

where $\boldsymbol{\eta}_k$ represents an additive noise, assumed to be Gaussian with zero-mean and σ^2 variance. Having K low resolution frames, $\mathbf{y}_k, k = 1, \dots, K$, we wish to obtain the high resolution image \mathbf{z} . To simplify notation, we will stack all the K available low resolution frames into a single column vector \mathbf{y} , and denote the length of this vector by M . Thus, M is the total number of low resolution pixels observed, in other words, if the k -th low resolution frame had M_k pixels, then $M = M_1 + \dots + M_K$.

Similarly, we will also stack the transformation matrices into a single matrix \mathbf{W} . This will then have $M \times N$ elements. Finally, the noise components will also be stacked into an M -dimensional column vector $\boldsymbol{\eta}$. Then, the observation model in a vectorised form may be written as:

$$\mathbf{y} = \mathbf{W} \mathbf{z} + \boldsymbol{\eta} \quad (2)$$

In this model, the unknown variable of interest is \mathbf{z} . The transformation matrix \mathbf{W} is usually parameterised, and as such, it is considered to be known up to a few parameters. Estimating these parameters of \mathbf{W} may be

done simultaneously with inferring \mathbf{z} . In this case we talk about the 'blind' super-resolution problem. A special case of this, when \mathbf{W} consists of blurring only, is often termed as 'blind deblurring' or 'blind deconvolution'. Several authors have tackled this problem with success.

In this work, we will consider the transformation matrix \mathbf{W} as being known, since the focus of our study is another aspect, namely the image prior, as it will become clear in the next section. However, inspired from new research in signal processing [11] that tries to exploit the good properties of certain random matrices in signal acquisition, we take our \mathbf{W} as a random matrix with entries drawn i.i.d. from a standard Gaussian and then fixed.

While the conventional transformation, \mathbf{W} is a product of blurring and down-sampling matrix of size $[M \times N]$, usually ill-conditioned matrix that models a linear blur operation and the down-sampling by row and column operator. This down-sampling operator made the problem harder where now we have less pixels to observe and wish to recover with a higher resolution image.

B. The Joint Model

Overall model is the joint model of the observations \mathbf{y} and the unknowns \mathbf{z} . That is, $Pr(\mathbf{y}, \mathbf{z})$. To assemble this from the previously presented components, we first rewrite the observation model given in Section II in the form of a probability distribution of the observations \mathbf{y} given the ground truth \mathbf{z} . That is, $Pr(\mathbf{y}|\mathbf{z})$. Using these, we have joint probability

$$Pr(\mathbf{y}, \mathbf{z}) = Pr(\mathbf{y}|\mathbf{z})Pr(\mathbf{z}) \quad (3)$$

where the first term is the observation model and the second term is the image prior model. Hence we have for the first term in (3):

$$Pr(\mathbf{y}|\mathbf{z}) \propto \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{W} \mathbf{z})^T (\mathbf{y} - \mathbf{W} \mathbf{z}) \right\} \quad (4)$$

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 (3) will be instantiated with either one of the image priors discussed in Section III. 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} . Again, this encodes knowledge about high resolution images in general, without any reference to the actual observed images \mathbf{y} . Recall that our task is to infer or estimate the high resolution image from its low resolution versions. To achieve this, now that we have formalised the problem, we need to 'invert' the causality described by our model, to infer the latent variables \mathbf{z} from the observed variables \mathbf{y} .

C. Inverting the model to estimate z

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

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

This is called the *posterior* probability of z given the observed data \mathbf{y} . Eq. (5) says that, the probability that z is the hidden image that gave rise to what we observed, i.e. \mathbf{y} , is proportional to the likelihood that this z fits the data \mathbf{y} and the probability that this bunch of N intensity values, i.e. the vector z , actually 'looks like' a valid image. Note that the latter is desperately needed in underdetermined systems, since there are infinitely many vectors z that fit the data.

D. Maximum A Posteriori Inference through Optimisation

To obtain the most probable estimate of z that conforms to our model and data, we need to maximise (5) as a function of z . Observe that, the denominator, $Pr(\mathbf{y})$ does not depend on z . Hence, the maximum value of the fraction (5) occurs for exactly the same z for which the maximum of the numerator does. That is, the most probable estimate is given by:

$$\hat{z} = \arg \max_z \frac{Pr(\mathbf{y}|z)Pr(z)}{Pr(\mathbf{y})} \quad (6)$$

$$= \arg \max_z Pr(\mathbf{y}|z)Pr(z) \quad (7)$$

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:

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

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:

$$Obj(z) = -\log[Pr(\mathbf{y}|z)] - \log[Pr(z)] \quad (9)$$

The most probable estimate is the \hat{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 (9), 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. (4):

$$l(z) := -\log Pr(\mathbf{y}|z) + const. \quad (10)$$

$$= \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{W}z)^T(\mathbf{y} - \mathbf{W}z) \quad (11)$$

III. 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:

$$Pr(z_i|z_{-i}) = Pr(z_i|z_{4\text{neighb}(i)}) \quad (12)$$

$$= Pr(z_i - \frac{1}{4} \sum_{j \in 4\text{neighb}(i)} z_j) \quad (13)$$

where the notation 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.

$$Pr(z) \propto \prod_{i=1}^N Pr(z_i|z_{4\text{neighb}(i)}) \quad (14)$$

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

The notation ' \propto ' 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 z . Since in this work we only need to estimate 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. [5], [6]. Alternatives include the so-called total variation model, employed e.g. in e.g. [8], which is based on image gradients, also quite simple. In [7], an experimental comparison of these two alternatives suggests these have comparable performance, the former being slightly superior though.

The simplicity of (15) 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 (15), $Pr(z_i - \frac{1}{4} \sum_{j \in 4\text{neighb}(i)} z_j)$, 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:

$$z_i - \frac{1}{4} \sum_{j \in 4\text{neighb}(i)} z_j = \sum_{j=1}^N d_{ij} z_j \quad (16)$$

$$= \mathbf{D}_i \mathbf{z} \quad (17)$$

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 neighbour-hood 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.

A. Gaussian-MRF

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

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

$$= \exp \left\{ -\frac{1}{2\lambda} \sum_{i=1}^N (\mathbf{D}_i \mathbf{z})^2 \right\} \quad (19)$$

where λ is the variance parameter.

B. 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

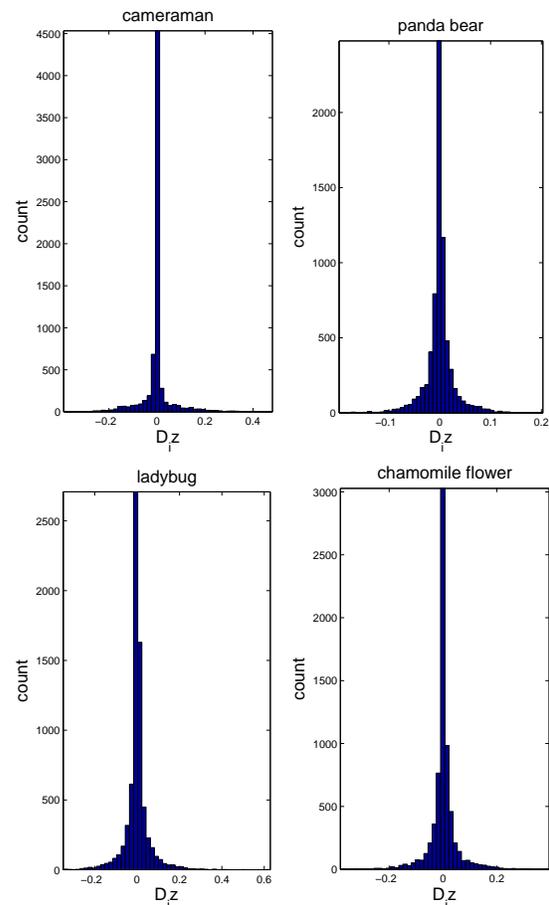


Fig. 1. Examples of histograms of the distribution of neighbourhood features $\mathbf{D}_i \mathbf{z}$, $i = 1, \dots, N$ from natural images.

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.

$$H(u|\delta) = \begin{cases} u^2, & \text{if } |u| < \delta \\ 2\delta|u| - \delta^2, & \text{otherwise.} \end{cases} \quad (20)$$

The Huber-MRF prior is then defined in (22) where λ is similar to a variance parameter.

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

$$= \exp \left\{ -\frac{1}{2\lambda} \sum_{i=1}^N H(\mathbf{D}_i \mathbf{z}|\delta) \right\} \quad (22)$$

IV. PEARSON TYPE VII-MRF

A. The univariate Pearson Type VII-MRF

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

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

where ν and λ control the shape of the distribution.

B. 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:

$$Pr(\mathbf{z}) \propto \left\{ \sum_{i=1}^N (\mathbf{D}_i\mathbf{z})^2 + \lambda \right\}^{-\left(\frac{\nu+N}{2}\right)} \quad (24)$$

C. Discussion on the two versions of Pearson-MRF

The version devised in Section IV-A 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 IV-B, in turn, does not allow such independence interpretation. Conversely, this can have 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.

V. EXPERIMENTS AND DISCUSSION

A. Experimental Setting

We present two sets of a single frame image super resolution experiments illustrating the performance of the hyper-parameters for testing the Pearson prior. We compare the state of the art image priors such as Gaussian and Huber. The LR image is blurred by the uniform blur matrix of size 3×3 , down-sampled by factor 4 and contaminated by standard deviation of Gaussian noise of 0.001, 0.01, 0.05 and 0.1. All images are in size $[100 \times 100]$ and the pixel intensities are scaled to interval $[-0.5, 0.5]$. The initial guess is initialized with Gaussian-MRF with σ^2/λ set to 1 and was used as a starting point for the recovery algorithm in previous work [3].

In this paper, we address the issue of parameter selection in [3] and improved it by estimating ν and λ . For this automated estimation, we initialised with a product of the inverse transformation matrix and the low resolution. We employed a conjugate gradient type method¹, which requires the gradient vector of the objectives.

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

Previously in paper [4], we applied the compressive matrix of \mathbf{W} to find out how well is the proposed image prior based MRF in comparison with the state of the art image priors, and hyper-parameters is manually tuned to acquire the optimum mean square error for all methods. We then observe how good the new approach of our automated parameter estimation for univariate Pearson type VII and other prior as well in under-determined problem.

B. Parameter Estimation Algorithm and Results

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 recovery. Since we are assessing the performance for both ν and λ , the recovery algorithm is assuming knowing the true noise variance σ^2 . From the observation in [3] using the constructed blur and down-sampling matrix \mathbf{W} , we found practical range of λ and ν . We made use this good range in our automated estimation to reduce the cost computation.

Recap, our parameter selection for ν and λ are found from the lowest mean square error from a several possible search space. Therefore, we overcome this issue by implementing hold out estimation and cross validation. Hold out and cross validation is a statistical method of evaluating and comparing learning algorithms by dividing data into two partitions: (i) one used to learn or train a model and the remainder used to validate the model [9]. Validation is done by estimating its minimum error of the mean squared error on how likely is the observed data, y with the model \mathbf{Wz} .

For compressive measurements, we develop hold out estimation in terms of reducing cost computation to recover the best solution. It is due of the variable itself that requires more computation to be done. Hence, it is sufficient to propose this method for random transformation. On the other hand, we implement k -folds cross validation for the classical transformation because the structured is sparse and this made the algorithm faster to be executed. To reduce variability, five rounds of cross-validation are performed using different folds, and the validation results are averaged over the rounds.

In k -fold cross validation technique, the data set is randomly partitioned into k groups. The learning algorithm is then trained k times, using all the training set data point except those in the k^{th} group. Both forms of the algorithms are described as follow in Algorithm 1 and 2. Indeed, in the approach described, the algorithm is less expensive and more precise search space is tested. Figure 2 presents the variation performance using proposed Algorithm 1. Then followed by Algorithm 2 which applied 5-folds for classical transformation. The performance of 3-dimensional among ν , λ and MSE are illustrated in Figure 3 and 4.

Algorithm 1 : Hold out estimation

- 1: **Goal:** To find optimal ν and λ by training a model using the training data set and finding the minimum error is found from the validation data set.
- 2: **Inputs:** training data, validation data, number of k -groups, ν and λ range, variance σ^2
- 3: **Outputs:** optimal ν , optimal λ , optimal error(MSE)
- 4: Randomize and divide data set into two groups: 5% for validation and the remainder is used for training set.
- 5: **for** $i = 1$ **to** $length(\nu)$ **do**
- 6: **for** $j = 1$ **to** $length(\lambda)$ **do**
- 7: Minimise with respect to \mathbf{z} using training set.
- 8: Compute performance (error): $mean((y(validate)-w(validate)*z(training))^2)$
- 9: Record the performance matrix error.
- 10: **end for**
- 11: **end for**
- 12: Find ν and λ that belong with the minimum error.
- 13: Minimise with respect to \mathbf{z} using the whole data set together with optimal ν and λ .

To compare the performance fairly, this hold out estimation is applied to all method image priors in this experiment. Here we test two different set of images and size to find out the effectiveness of Pearson prior in those two cases. In general, it is well known that no best image prior can best fit on every data. Nevertheless, these results in Figure 3 do demonstrate that our method is competitive with the state of the art on that type of data when other image priors are estimated automatically too.

Note that the loops need not completely converge. It is sufficient to increase and not necessarily minimise the objective at each combination. Nevertheless, we observed the final minimisation is converge faster by letting more iterations once the algorithm used the optimal value of ν and λ . Next, we estimate parameter using k -folds cross validation as described in Algorithm 2 for the conventional and complicated transformation matrix W .

Our proposed algorithm for classical transformation illustrates the performance result over 5-folds cross validation in Figure 4. All the competing image priors used this automated estimation and the comparison is to find out how good is the Pearson prior when the parameter estimation is no longer chosen by the best manual selection as presented work in [3]. 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 and give a competitive solution to Huber-MRF across the four levels of noise.

We also observe how does this automated hyper-parameter estimation of our Pearson type VII based MRF prior compare to these manual selections best results. Figure 6 shows the best manual results with reference to the ground

Algorithm 2 : k -fold cross validation for estimating ν and λ

- 1: **Goal:** To find optimal ν and λ by training a model using the training data set and finding the minimum error is found from the 5-folds cross validation.
- 2: **Inputs:** training data, validation data, number of k -groups, ν and λ range, variance σ^2
- 3: **Outputs:** optimal ν , optimal λ , optimal error
- 4: Randomize and divide data set into k -groups.
- 5: **for** $k = 1$ **to** $k - groups$ **do**
- 6: validate = find(group== k)
- 7: training = find(group≠ k)
- 8: **for** $i = 1$ **to** $length(\nu)$ **do**
- 9: **for** $j = 1$ **to** $length(\lambda)$ **do**
- 10: Minimize with respect to \mathbf{z} using training set.
- 11: Compute the performance found using the k -th set: $mean((y(validate)-w(validate)*z(training))^2)$
- 12: Record the performance matrix error.
- 13: **end for**
- 14: **end for**
- 15: Report the mean error over all k test sets.
- 16: **end for**
- 17: Find ν and λ that belongs to the minimum 5-folds error value.
- 18: Minimize with respect to \mathbf{z} using the whole data set based on the optimal ν and λ found.

truth image and the proposed Algorithm 2. The significance result indicates that this proposed algorithm performs well without gaining access to the true image when evaluating the performance image. Finally, Figure 7 presents the image recovery of super resolution using Pearson type VII based MRF.

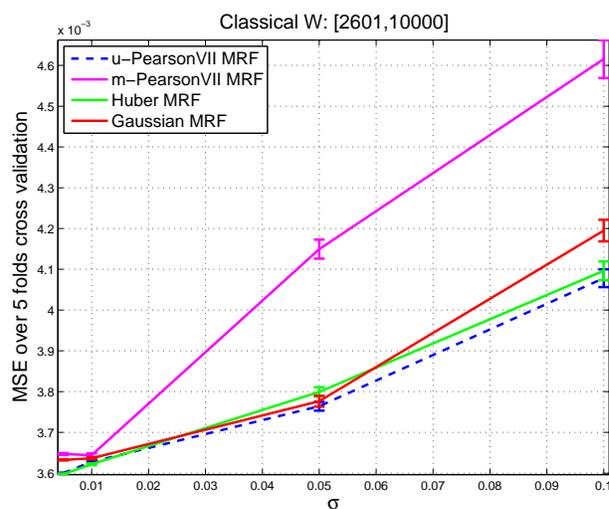


Fig. 5. Comparative MSE performance for under-determined system using cameraman image varying four levels of noise. The best value for hyper-parameters for every image prior found using 5-folds cross validation (cv) technique. The error bars are over 10 independent trials. Pearson prior maintains its good performance for every level of noise in terms of mean square error (MSE).

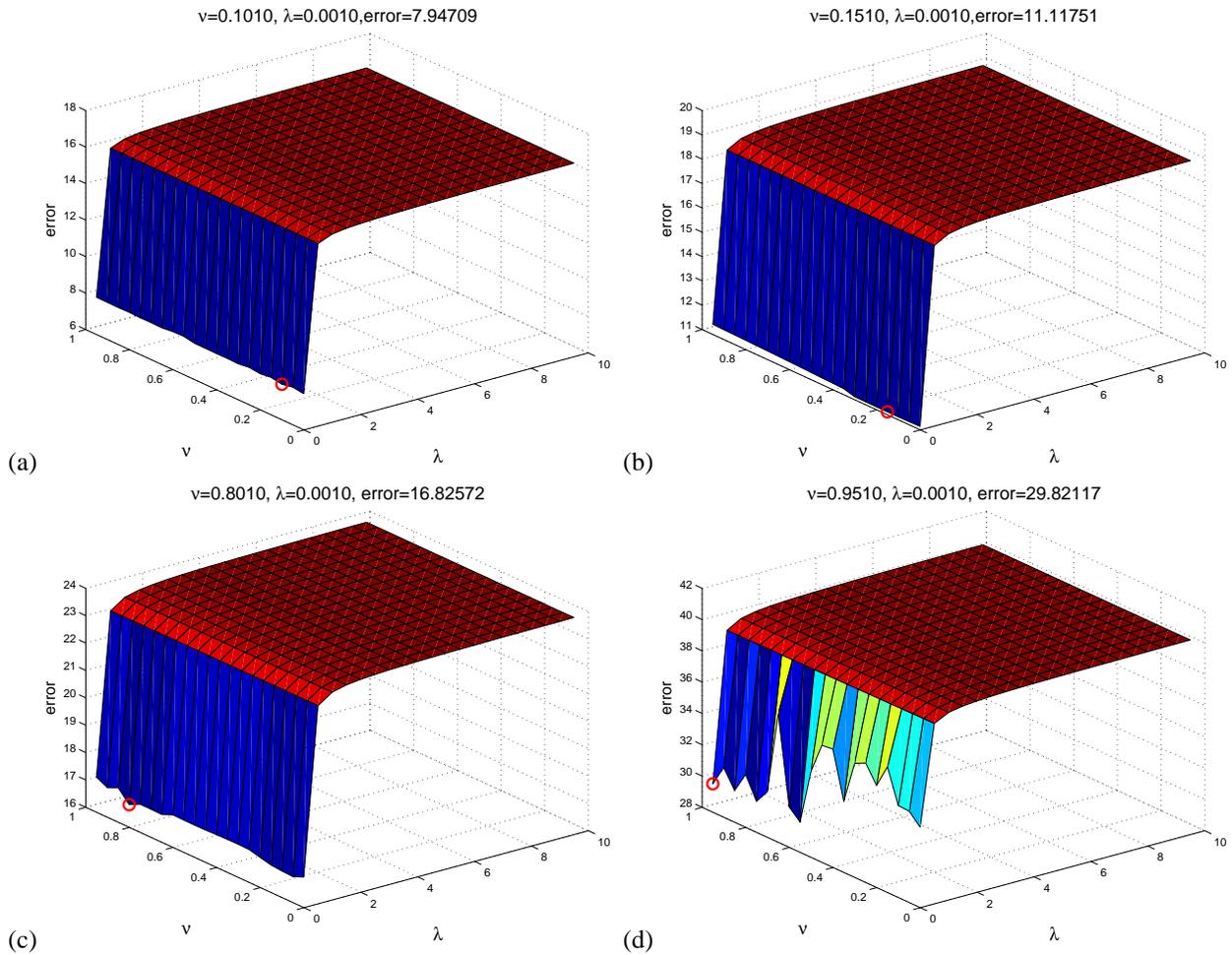


Fig. 2. 3-dimensional plot varying ν , λ and its mean squared error with variance: (a) 0.005, (b) 0.01, (c) 0.05 and (d) 0.1 using random transformation for data generation using cameraman image. Smaller noise shows a stable performance while higher noise performed inconsistently. However, both optimal values are found in a smaller range. We see the error performance is increasing rapidly when λ is searched from range 1 and its reaching the stability performance.

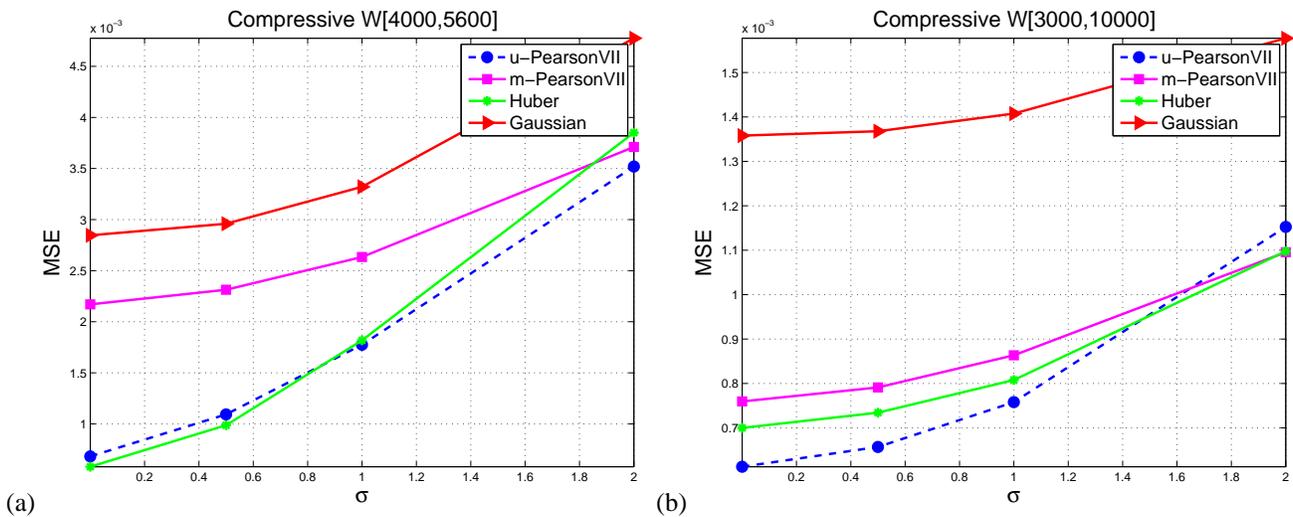


Fig. 3. Comparative MSE performance for under-determined system for two different images and size (a) synthetic data, (b) real data varying four level of noise using the best values of hyper-parameter for every image prior found using held out estimation. Once repetition shows that Pearson prior is superior in higher noise(right), however it is vice versa for $W[3000,10000]$. For lower noise, Pearson prior achieves its best performance when having less observation data in (b).

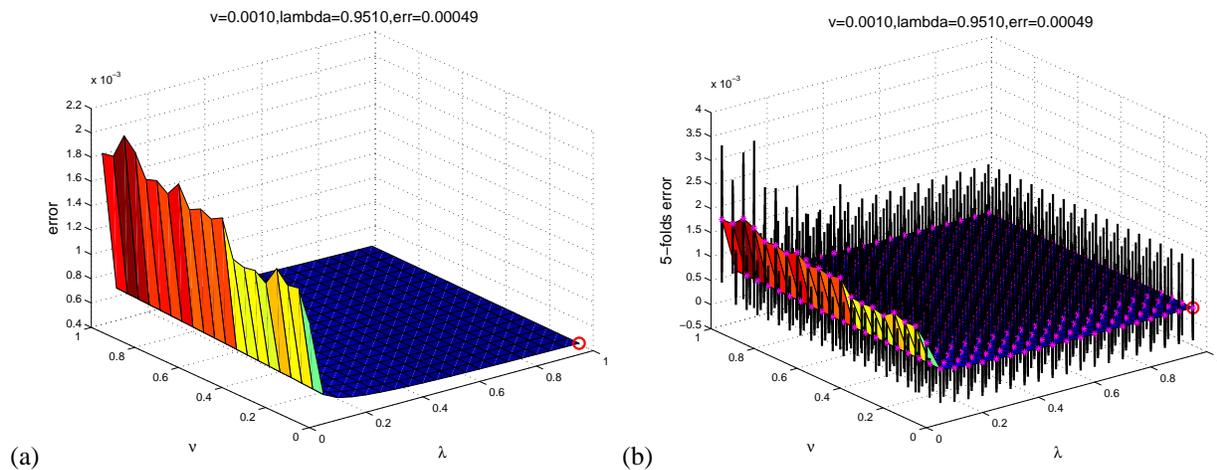


Fig. 4. Example of mean error over all k test sets (left) and mean and standard deviation over 5-folds repetition (right) for variance, 0.005 using transformation matrix of blur and down-sampled version.

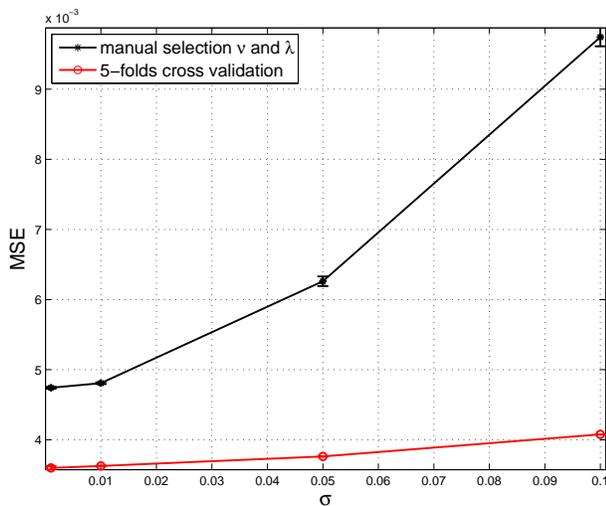


Fig. 6. Comparing the MSE performance of the fully automated Pearson type VII based MRF approach with the best MSE found by manual selection of the hyper-parameters in previous work [3]. The error bars are over 10 independent trials where additive noise and the transformation W was blurred and down-sampled. Note that the number of observation for manual selection is differ from automated estimation. W size for manual selection previously used is [2500,10000] while the automated estimation using 5-folds cross validation has [2601,10000]. Despite of having less number of measurements, our proposed estimation still perform better than the one with manual tuning with one of the parameter is fixed.

VI. CONCLUSION

Compressive measurement and classical super resolution has been considered from a probabilistic model based framework. We tested this on both synthetic data and real data in under-determined system. In this paper we formulated two versions of Pearson-MRF image priors, and conducted a comparative experimental study between 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 competitive with Huber-MRF in terms of qualitative measurement mean square error. Our proposed algorithm

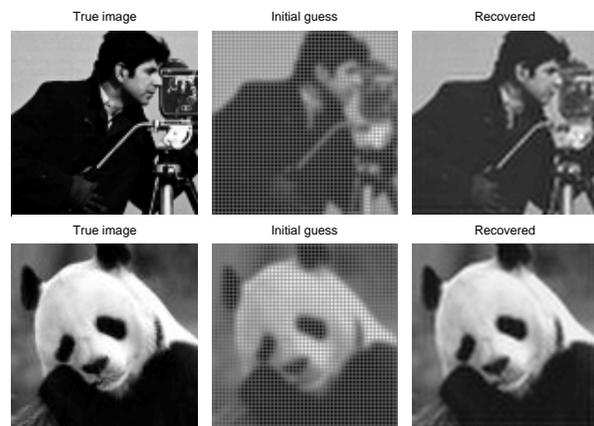


Fig. 7. Example image recovery of 'cameraman' (10000 pixels) from blurred and down-sampled to 2601 pixels and additive noise with $\sigma^2 = 1e-3$

for parameter estimation is conceptually simple, automated and easy to implement. The recovered image is always consistent although it has several local optima and we asses two set of 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. Future work is aimed towards recovering several images from multiple scenes for under-determined system too.

ACKNOWLEDGMENT

The author would like to express her gratitude to School of Computer Science, University of Birmingham, United Kingdom and Universiti Sains Islam Malaysia (USIM) for the support and facilities provided.

REFERENCES

- [1] J.Sun, A.Kabán and J.Garibaldi, *Robust Mixture Modeling using the Pearson Type VII Distribution*, *Pro. Int. Joint Conference on Neural Network*, 2010
- [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] Sakinah Ali Pitchay, *Non-Linear Restoration from a Single Frame Super Resolution Using Pearson Type VII Density*, *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering and Computer Science 2010, WCECS 2010*, 20-22 October, 2010, San Francisco, USA, pp426-431.
- [4] 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)*
- [5] 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.
- [6] 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.
- [7] 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.
- [8] L. C. Pickup, D. P. Capel, S. J. Roberts, A. Zissermann, *Bayesian Methods for Image Super-Resolution*, *The Computer Journal*, 2007.
- [9] Payam Refaeilzadeh, Lei Tang and Huan Liu. *Cross-Validation*, Arizona State University, November 2008.
- [10] Emmanuel Candes and Justion Romberg, Caltech. *l_1 -MAGIC: Recovery of Sparse Signals via Convex Programming* October 2005.
- [11] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk *it Signal Processing with Compressive Measurements*, *IEEE Journal of Selected Topics in Signal Processing*, 2009.