Analysis of Multi-frame Super Resolution Methods

Evan Widloski
UIUC
evanw3@illinois.edu

Akshayaa Magesh
UIUC
amagesh2@illinois.edu

Aditya Deshmukh
UIUC
ad11@illinois.edu

1. Introduction

Super-resolution (SR) is the process of obtaining high resolution (HR) images from one or more low resolution (LR) observations. Multi-frame super-resolution refers to the case where multiple images of the same scene are available. In this project, we address the case of multi-frame super resolution [3]. The degradation in the low resolution images could be a result of one or more of the following reasons: camera or scene motion, camera zoom, focus and blur. Super resolution is possible because each low resolution image contains pixels that represent subtly different functions of the original image. Note that super resolution is different from techniques such as interpolation, restoration and image rendering. In super-resolution, besides improving the quality of the output, its size (number of pixels per unit area) is also increased.

Super-resolution finds applications in satellite and aerial imaging, medical image processing, facial image analysis, biometrics recognition, text image analysis, sign and number plates reading, to name a few.

In this project, we propose to implement and evaluate a handful of super-resolution methods which are either considered state of the art, or are broadly representative of one of the many classes of super-resolution algorithms. We will evaluate their effectiveness by testing on a set of synthetically generated LR images and comparing the reconstruction to the original HR image.

2. Details of the Approach

The following represents the most common forward model that captures the generation process of degraded low-resolution images $\{y_k(m,n)\}_k$ from a high resolution image $x(u,v)$:

$$y_k(m,n) = d(b_k(w_k(z(u,v)))) + \eta(m,n)$$

where k denotes the index of the degraded low-resolution image, and d, b_k, w_k are downsampling, blurring, and warping operators respectively. This model is visualized in Figure **??**. In this project, we restrict the warping to be and rotation. The forward model is discussed in detail in 2.1.

There are two main parts to obtaining a HR image from multiple LR images. The first step is image registration. The purpose of this step is to obtain the registration parameters of the forward model such as the translation and rotation shifts across the frames with respect to some fixed high resolution grid. In this project, we implement a multi-frame subpixel registration algorithm described in Subsection 2.2.

The second step is the reconstruction of the HR image from the LR images using a suitable estimation framework. In this project, we look at two reconstruction mechanisms, a joint MAP registration and estimation method [2], and a MAP estimation using Huber Markov Random Fields model [4]. The joint MAP registration and estimation method is discussed in detail in Subsection **??** and the MAP estimation using Huber Markov Random Fields is described in Subsection 2.3

2.1. Forward Model

In this section, we describe the assumptions and approach used to implement the forward model. This model accounts for physical processes involved while observations are being made, such as optical distortion from the imaging equipment (lens aberrations, atmospheric distortion, etc.) motion blur from relative movement of the sensor and the scene (shaking hand holding a camcorder, spacecraft orbital motion) and a sampling operation due to the discrete nature of imaging sensors.

In this project, we constrain the motion between frames to be constant, linear and purely translational. This is a useful model for many remote sensing applications where relative motion between an imaging satellite's and the ground is well approximated by translation in the field of view.

For computation, we begin by assuming we have access to some input $x(u,v)$ known as the *high-resolution image* which is unperturbed by the processes mentioned above. This scene might come from a simulated ground truth, or could be obtained as an output from a super-resolution algorithm. The input scene is assumed to be spatially discrete on a grid known as the *high-resolution grid*. The choice of this grid resolution has implications on the simulation fidelity and complexity. A choice of high resolution gives a

higher simulation fidelity at the expense of computational complexity and vice-versa.

There are also *low-resolution images* which exist on a *low-resolution grid*. These low-resolution images correspond to noisy, blurred observations made by the imaging system and their size is determined by the parameters of the imaging sensor. The high and low-resolution are related by the *downsample factor* as shown in Figure 1. This factor relates the ratio of scale of objects as they appear in the high-resolution image and observations. For example, if the high-resolution image has pixels which correspond to a 4m by 4m patch in the physical scene, and the low-resolution pixels correspond to 16m by 16m, the downsample factor is 4.

There are three steps to obtain the low-resolution images given a high-resolution image.

First we compute the motion blur kernel. This is a 1 pixel wide line which traces out the path of the sensor while a frame is being captured. The length of exposure and frame rate of a sensor control the length of this line.

Next, the downsample factor is accounted for by convolving the single pixel path with a square of size equal to downsample factor.

A final convolution with a blurring kernel accounts for any effects of the optical path.

These three steps can be combined to form a single observation kernel which is applied to all pixels, as shown in Figure 3.

With the observation kernel in hand, we can now form the low-resolution images by integrating along the path and applying the blurring kernel, as in Figure 2.

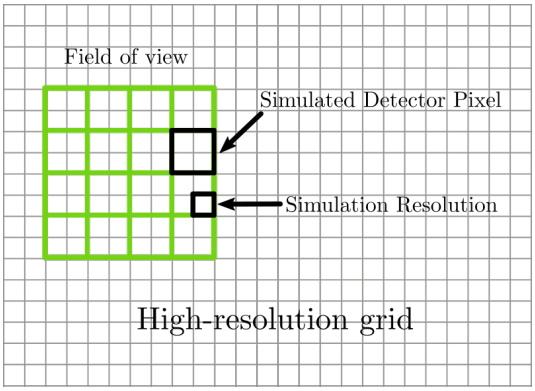


Figure 1. Visualization of high and low-resolution grids. In this case, downsample factor is 2.

2.2. Multi-frame Registration Algorithm

In this section, we describe a multi-frame subpixel registration algorithm which has been developed from a paper by Guizar-Sicairos [1]. This algorithm is incorporated in the super-resolution methods in later sections.

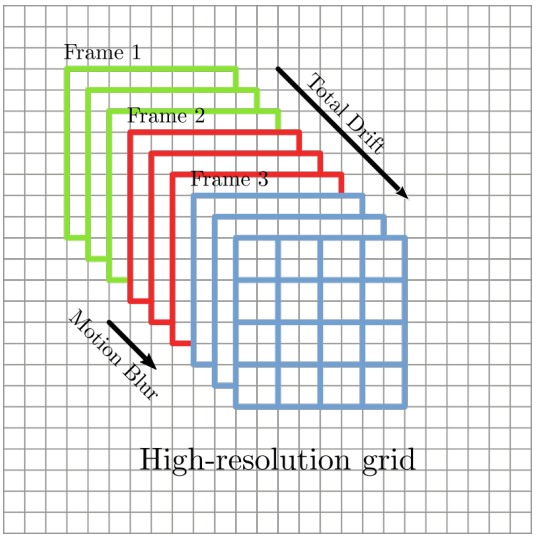


Figure 2. Integration path of individual frames for linear motion.

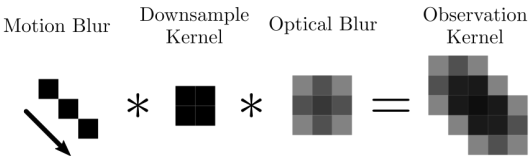


Figure 3. We can convolve the kernels from the motion blur, downsample and optical blurring steps to form a single observation kernel.

Guizar-Sicairos [1] describes a two-frame subpixel registration method which is designed for simple translation motion. It is a 2-step subpixel registration method which first obtains a coarse, whole-pixel drift vector estimate from standard cross-correlation and then refines it by upsampling the cross-correlation with sinc interpolation.

This is described mathematically:

$$\text{coarse} = \text{argmax}_{\boldsymbol{x}\text{IFFT}\left[Y_i \cdot \overline{Y_j}\right](\boldsymbol{x})}$$

$$\text{fine} = \text{argmax}_{\boldsymbol{x}\in\text{N(coarse)}}\text{UpsampIDFT}\left[Y_i \cdot \overline{Y_j}\right](\boldsymbol{x})$$

where Y_i and Y_j are the Fourier transforms of the low-resolution images being registered, $\cdot$ is elementwise product, and $N(coarse)$ is a small neighborhood of pixels around the coarse estimate. These coarse and fine estimates are then fused to obtain a complete subpixel registration.

A visualization of this process is shown in Figure 5.

The major argument of this paper is that the refined estimation step has been greatly accelerated by directly evaluating the IDFT in a region around the coarse estimate as opposed to a zero-padded IFFT technique.

The multi-frame registration method we develop in this paper takes a similar two-step approach. It is trivial to extend this two-frame registration method to multiple frames by simply computing a registration estimate for all adjacent

frames and taking the average. However, this approach performs poorly in high-noise setting when the peak of the cross-correlation is hidden in the noise floor. Instead, we propose to extend the Guizar-Sicairos method nontrivially by simultaneously using all low-resolution frames to compute a registration estimate once.

First we begin by substituting the circular cross-correlation above (implemented with FFTs) with an analog called the *phase correlation*. Phase correlation is similarly fast as circular cross-correlation, but has the benefit of a sharper peak corresponding to drift, as shown in 6.

$$PC_{i,j}(\boldsymbol{x}) = \text{IFFT}\left[\frac{Y_i \cdot \overline{Y_j}}{|Y_i \cdot \overline{Y_j}|}\right](\boldsymbol{x}) \quad = \delta(\boldsymbol{x} - \boldsymbol{v}(j-i))$$

where $\boldsymbol{v}$ is the interframe drift.

we then fuse these phase-correlations into a single image which we call a *correlation sum*.

$$CS_1 = PC_{1,2} + PC_{2,3} + PC_{3,4} + ...$$

This is shown in Figure 7.

We can similarly repeat this process for frames separated by 2, 3, 4, ...

$$CS_2 = PC_{1,3} + PC_{2,4} + PC_{3,5} + ...$$
$$CS_3 = PC_{1,4} + PC_{2,5} + PC_{3,6} + ...$$
$$...$$

We end up with a series of correlation sums as shown in Figure 4. The key observation is that the position of the peak of the correlation sum is vk, and since k is known, we can rescale each correlation sum so that all their peaks are located at v before being summed.

$$estimate = \arg\max_x \left[\sum_{k=1}^{K} \text{Upscale}_{K/k}(CS_k)\right](x)$$

where Upscale is an image upscaling operation.

2.3. MAP Esimtation using Huber MRFs

A lot of probabilistic approaches have been proposed for super-resolution from LR images. Mostly these approaches follow the method of MAP estimation, which surpasses ML estimation when the number of LR images is less. If the number of the LR images is insufficient for the determination of the super-resolved image, the involvement of a-priori knowledge plays an important role and MAP outperforms ML.

Markov Random Fields (MRFs) have been widely used in the literature for modeling a-priori knowledge, since they exhibit the ability to capture the notion of similarity between neighboring pixels. The Huber-Markvov Random Field(HMRF) models a Gibbs prior on the HR images:

$$P(\boldsymbol{z}) = 1/Z \exp\{-1/2\beta \sum_{c \in \mathcal{C}} \rho_\alpha(\boldsymbol{d}_c^t \boldsymbol{z})\}. \tag{1}$$

The discussion on above quantities is beyond the scope of this report, however the motivation to consider the function

$$\rho_\alpha(x) = \begin{cases} x^2, & |x| \le \alpha \\ 2\alpha|x| - \alpha^2, & |x| > \alpha \end{cases} \tag{2}$$

is that it preserves edges by penalizing less severely than other methods like Tikhonov regularization. The optimization function for MAP estimation can then be written as:

$$f(\boldsymbol{z}, \alpha, \Lambda) = ||\Lambda^{1/2}(\boldsymbol{y} - A\boldsymbol{z})||^2 + \sum_{m,n,r} \rho_\alpha(\boldsymbol{d}_{m,n,r}^t \boldsymbol{z}). \tag{3}$$

The algorithm adopted in [4] is gradient projection to minimize this objective function. The algorithm fixes an LR image as the anchor and tries to find the HR image associated with it by describing the other LR images as linear transformations of the HR image. The restriction that the anchor LR image is solely generated by the HR images puts a constraint on the HR image to be reconstructed. This constraint is then used in the gradient descent algorithm to project the gradient in the appropriate direction.

Here are a few observations we noted:

1. The initialization is a very crucial step and the we found that initializing using bicubic interpolation gave better and faster results than the proposed initialization which involves multiplication of anchor LR transformation matrix A_0^T with the anchor LR image y_0.

2. The step size for gradient descent involves computation of the Hessian matrix. This is extremely tedious and we replaced it with small step sizes with large number of iterations. We also tried implementing avaliable optimizers like 'L-BFGS' and 'Netwon-CG', but owing to the large dimensions, these didn't work.

3. The accuracy of reconstruction also depended on the hyper-parameters α and β (which controls Λ), and a good tradeoff is necessary to optimize the objective function. We found that setting $\alpha = 10$ and $\beta = 1$ worked the best for the dataset considered.

4. ML solution for the considered experiment did not work since the matrix $A^T A$ was singular.

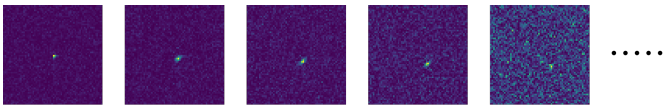


Figure 4. Correlation sums. The peak location is a function of the interframe drift and the correlation sum index

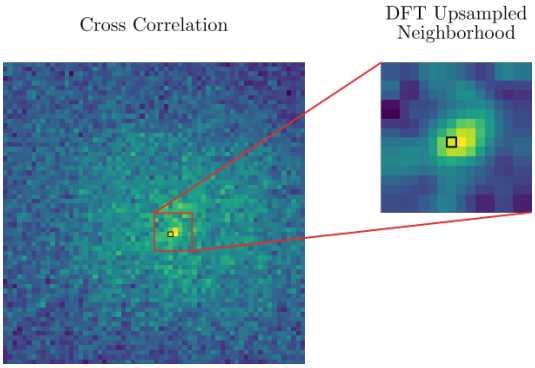


Figure 5. Visualization of two-frame registration method described in [1]. The small black squares correspond to the coarse and fine drift estimates.

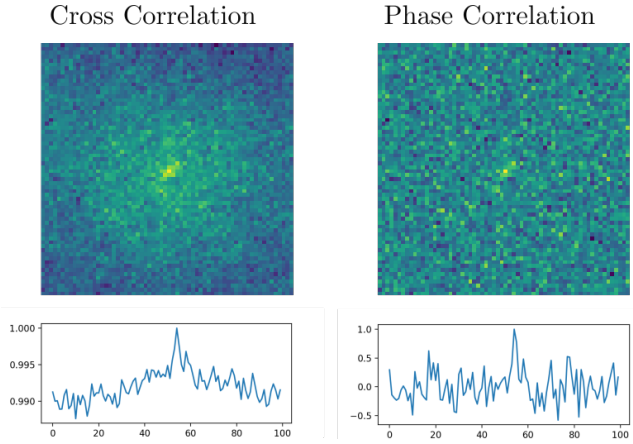


Figure 6. Comparison of cross-correlation and phase correlation. Phase correlation has a more sharply defined peak with a uniform noise-floor.

2.4. Joint MAP registration and Estimation

The model adopted in [2] is as follows:

$$y_{k,m} = \sum_{r=1}^{N} w_{k,m,r}(\boldsymbol{s}_k) z_r + \eta_{k,m}, \tag{4}$$

where $w_{k,m,r}(\boldsymbol{s}_k)$ denotes the contribution of the rth high-resolution pixel to the mth low resolution observed pixel in the kth frame. The total number of pixels in the high resolution image is N. There are p frames and the size of the low resolution image is M. The vector $\boldsymbol{s}_k$ contains the K registration parameters for frame k. These parameters capture the notions of global translation and rotation. The weighted sum models the blurring of the underlying scene values due to the finite detector size and the PSF of the optics.The term $\eta_{k,m}$ represents additive noise, that will be assumed to be i.i.d Gaussian with variance σ_n^2. A Gaussian noise model is useful for a variety of imaging systems and scenarios. In most practical imaging applications, detected photons follow Poisson statistics. The model can be written in matrix form as:

$$\mathbf{y} = \mathbf{W_s z} + \mathbf{n} \tag{5}$$

The observation model in (4) assumes that the underlying scene samples, z, remain constant during the acquisition of the multiple low resolution frames, except for any motion allowed by the motion model.

In many practical imaging situations, the registration parameters are not known a priori. Therefore, one way to handle this is to consider them to be random parameters to be estimated along with the high resolution image z, as done in [2]. Another way to approach this would be to use the multi-frame registration method described in Subsection 2.2 to obtain estimates if $\boldsymbol{s}_k$. In our application, we consider the low resolution images to be frames in a video sequence with a fixed frame rate, drift velocity and drift angle. Thus, there are two registration parameters we consider in this case, the drift velocity and drift angle.

In the first method, the MAP estimates are computed as:

$$\hat{\mathbf{z}}, \hat{\mathbf{s}} = \arg\max_{\mathbf{z},\mathbf{s}} P(\mathbf{z}, \mathbf{s}|\mathbf{y}) \tag{6}$$

$$= \arg\max_{\mathbf{z},\mathbf{s}} P(\mathbf{y}|\mathbf{z}, \mathbf{s}) P(\mathbf{z}) P(\mathbf{s}) \tag{7}$$

The prior $P(\mathbf{z})$ is chosen as a Gibbs prior where the exponential term is a sum of clique potential functions as follows:

$$P(\mathbf{z}) = \frac{1}{(2\pi)^{N/2}|C_z|^{1/2}} \exp\left(-\frac{1}{2\lambda}\sum_{i=1}^{N}\left(\sum_{i=1}^{N} d_{i,j} z_j\right)^2\right) \tag{8}$$

The coefficient vectors $\mathbf{d}_i$ for $i = 1, ..., N$ effectively express a priori assumptions about the local relationships between pixel values in $\mathbf{z}$. These parameters are selected to provide a higher probability for smooth random fields. The following parameters have been selected for the coefficient

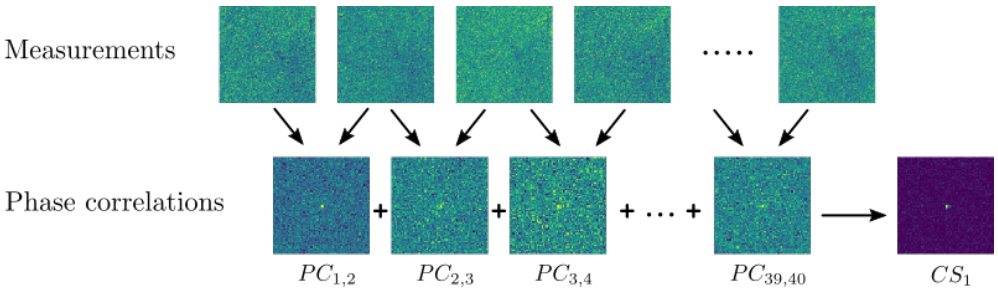


Figure 7. Visualization of computing correlation-sum

parameters:

$$d_{i,j} = \begin{cases} 1, & \text{for } i = j \\ -1/4 & \text{for } j : j\text{is cardinal neighbor of } i \end{cases} \quad (9)$$

The parameter λ is a tuning parameter, can be used empirically to control the penalty for discontinuities and rapidly varying features in $\mathbf{z}$.

If the signal-to-noise ratio is high and there are a relatively small number of registration parameters to estimate, a no preference prior can yield a useful solution. Thus, the estimation of the registration parameters reduces to a maximum likelihood estimate.

The method proposed in [2] proposes a cyclic coordinate-descent optimization technique to get an estimate of the registration parameters given the current estimate of the HR image by performing a search over the possible registration parameters. The loss function forms a quadratic function in $\mathbf{z}$ and can be minimized readily with respect to $\mathbf{z}$ if the registration information $\mathbf{s}$ is fixed.

In order to get the estimates for the registration parameters at each iteration, we initially tried a grid search over all possible registration parameters. However, this was computationally expensive to calculate the forward model. Thus, this step from [2] is replaced by the multi-frame registration method proposed in Subsection 2.2, which is much faster.

The optimization procedure is initialized using an initial estimate of the high resolution image obtained by a cubic interpolation of the first LR image frame. The gradient and the optimal step for the gradient descent optimization to update the estimate of the high resolution image can be computed very efficiently through the use of convolution and use of the forward model. Thus, combining the gradient descent proposed in this paper, with the registration step proposed in Subsection 2.2 gives good results in very less time.

3. Results

We performed experiments on a synthetic dataset, the Mars Orbiter dataset that captured the images of the Martian surface. 40 LR images were generated from a fixed scene with a frame rate of 4 Hz. The drift velocity of the the frames is 1 pix/s and the drift angle is 45 deg. The high-resolution image is of size 500 by 500 and the low-resolution images are of size 100 by 100. The downsampling factor considered is 4. The value of the tuning parameter λ for the joint MAP registration and estimation (Subsection 2.3) is 100.

3.1. Registration Error

The registration algorithm performs surprisingly well at low SNRs. As seen in Figure 7 (-10dB SNR), the high levels of noise in the observations complete obscure any salient features in the image, and yet we are still able to achieve subpixel levels of registration. To further test the limits of the algorithm, we have run a Monte Carlo simulation shown in 8. We sweep the SNR and plot the mean registration error. The shaded regions correspond to +/- 1 standard deviation.

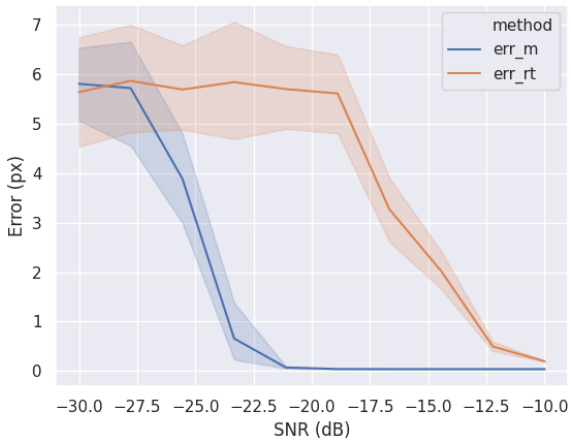


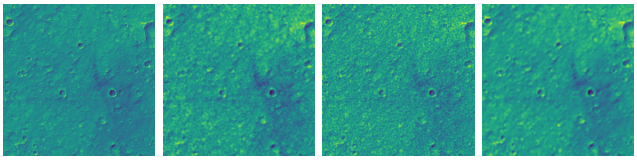
Figure 8. Naive registration vs our multiframe algorithm. This experiment was run with 40 video frames of size 100 by 100. Our registration algorithm begins to fail at around -25dB.

	Shift-and-add	HMRF	Joint MAP
PSNR	17.08	16.35	22.03
SSIM	0.51	0.28	0.50

Table 1. PSNR and SSIM metrics

3.2. Reconstructions

We implemented three algorithms for super-resolution reconstruction: a simple upsample-shift-and-add algorithm, MAP reconstruction using HMRFs and Joint MAP registration and reconstruction algorithm (as discussed in previous sections). We evaluated the reconstructions using peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [5] metrics. Figure 3.2 shows the reconstructions along with the original image. We observed that the joint MAP registration and reconstruction method worked the best, while MAP reconstruction using HMRF performed poorly. Table 1 gives the PSNR and SSIM metric evaluated on the reconstructions.



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

Figure 9. (a) Original (b) Upsampled and coadded reconstruction (c) Joint MAP super-resolution reconstruction (e) HMRF super-resolution reconstruction

4. Discussions and Conclusions

In this paper, we implemented a multi-frame registration method developed from a simpler two-frame method [1]. We have implemented and compared the performance of 3 reconstruction algorithms: a naive shift and add method, a MAP estimation method using HMRF prior model and a joint MAP registration and estimation procedure. We have the final reconstructed images obtained by using these methods and have evaluated the performance of these methods using PSNR and SSIM metrics. Both the MAP estimation using HRMF prior models and the joint MAP registration and reconstruction methods have a computation time of less than a few minutes for the dataset used in our experiments.

5. Statement of Individual Contribution

- Evan Widloski: Implemented forward model and developed fast multiframe registration algorithm. Implemented rudimentary super-resolution via coadding registered frames to compare against MAP and ML super-resolution methods.
- Akshayaa Magesh: Studied and analyzed the Maximum aposteriori (MAP) methods of reconstructing a HR image from multiple LR images using methods inspired from [2] and [4]
- Aditya Deshmukh: Studied and analyzed the Maximum aposteriori (MAP) methods of reconstructing a HR image from multiple LR images using methods inspired from [2] and [4]
- Both code and video can be found online here: https://github.com/evidlo/sr549

References

[1] S. T. Fienup J. R. Guizar-Sicairos, M. Thurman. Efficient subpixel image registration algorithms. *Optics Letters Volume 33*, 2008. 2, 4, 6

[2] Russell C Hardie, Kenneth J Barnard, and Ernest E Armstrong. Joint map registration and high-resolution image estimation using a sequence of undersampled images. *IEEE transactions on Image Processing*, 6(12):1621–1633, 1997. 1, 4, 5, 6

[3] T Nasrollahi, K. Moeslund. Super-resolution: a comprehensive survey. *Machine Vision and Applications Volume 25*, 2014. 1

[4] Richard R Schultz and Robert L Stevenson. Extraction of high-resolution frames from video sequences. *IEEE transactions on image processing*, 5(6):996–1011, 1996. 1, 3, 6

[5] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004. 6