Super-Resolution Imaging : Use of Zoom as a Cue

M.V. Joshi and Subhasis Chaudhuri
Department of Electrical Engineering
Indian Institute of Technology - Bombay
Powai, Mumbai-400076. India
{mvjoshi, sc}@ee.iitb.ac.in

Abstract

In this paper we propose a novel technique for super-resolution imaging of a scene from observations at different zooms. Given a sequence of images with different zoom factors of a static scene, the problem is to obtain a picture of the entire scene at a resolution corresponding to the most zoomed image in the scene. We model the super-resolution image as a Markov random field (MRF) and a maximum a posteriori *estimation method is used to derive a cost function which is then optimized to recover the high resolution field. Since there is no relative motion between the scene and the camera, as is the case with most of the super-resolution techniques, we do away with the correspondence problem.*

1. Introduction

Availability of high spatial resolution images is often desirable in most computer vision applications. Be it remote sensing, medical imaging, robot vision, industrial inspection or video enhancement (to name a few), operating on high resolution images leads to a better analysis in the form of lesser misclassification, better fault detection, more true-positives, etc. However, acquisition of high resolution images is severely constrained by the drawbacks of sensors that are commercially readily available. Thus, images acquired through such sensors suffer from aliasing and blurring. Aliasing occurs as a consequence of insufficient density of the detector array which causes sampling of the scene at less than Nyquist rate, while blurring occurs due to integration of the sensor point spread function (PSF) at the sensor surface. Hence, one must resort to image processing methods to construct a high resolution image from one or more available low resolution images. *Super-resolution* refers to the process of producing a high spatial resolution image from several low resolution images. It includes upsampling the image thereby increasing the maximum spatial frequency that can be represented and removing degradations that arise during image capture, viz., aliasing and blurring. The effect of aliasing differs with zooming. Thus one can use zoom as cue for generating high resolution images at the lesser zoomed area of a scene.

Now we review some of the prior works on super-resolution imaging. Many researchers have tackled the super-resolution problem for both still and video images, e.g., [8, 9, 14, 18] (see [16] for details). Tsai and Huang [18] were the first to propose a frequency domain approach to reconstruction of a high resolution image from a sequence of undersampled low resolution, noise-free images. Kim *et al.* discuss a recursive algorithm, also in the frequency domain, for the restoration of super-resolution images from noisy and blurred images [10]. A minimum mean squared error approach for multiple image restoration, followed by interpolation of the restored images into a single high resolution image is presented in [15]. Ur and Gross use the Papoulis-Brown generalized sampling theorem to obtain an improved resolution picture from an ensemble of spatially shifted pictures [19]. However, these shifts are assumed to be known by the authors. An iterative backprojection method is used in [9], wherein a guess of the high resolution output image is updated according to the error between the observed and the low resolution images obtained by simulating the imaging process. But back-projection methods can be used only for those blurring processes for which such an operator can be calculated. A projection onto convex sets (POCS)-based method is described in [17]. A MAP estimator with Huber-MRF prior is described by Schultz and Stevenson in [13]. Elad and Feuer [5] propose a unified methodology for super-resolution restoration from several geometrically warped, blurred, noisy and downsampled measured images by combining ML, MAP and POCS approaches. An adaptive filtering approach to super-resolution restoration is described by the same authors in [6]. Chiang and Boult [3] use edge models and a local blur estimate to develop an edge-based super-resolution algorithm. Other approaches include MRF based super-resolution proposed by Rajan and Chaudhuri [12]. Here authors consider availability of decimated, blurred and noisy versions of a high resolution image which are used to generate a super-resolved image. A known blur acts as a cue in generating super-resolution image. For super-resolution applications they also propose a general-

ized interpolation method [11]. Here a space containing the original function is decomposed into appropriate subspaces. These subspaces are chosen so that rescaling operation preserves properties of the original function. On combining these rescaled subfunctions, they get back the original space containing the zoomed function. Shekarforoush *et al.* use MRFs to model the images and obtain a 3D high resolution visual information (albedo and depth) from a sequence of displaced low resolution images [8]. The effect of sampling a scene at a higher rate is acquired by having interframe subpixel displacements.Freeman *et al.* in [20] describe image interpolation algorithms which use a database of training images to create plausible high frequency details in zoomed images. In [1] authors develop super-resolution algorithm and called it as *hallucination* which uses information contained in a collection of recognition decisions in addition to the reconstruction constraints.

In this paper we present a new approach to generation of super-resolution intensity map using the zoom as a cue. We consider the resolution at which the most zoomed observed image is available as the resolution to which the entire scene needs to be super-resolved. This requires that we recover high resolution intensity values for the lesser zoomed entire area of a scene. Our approach generates a super-resolution image of the entire scene although only a part of the observed zoomed image has multiple observations. Our method does a noise smoothing for the region where multiple observations are available and utilizes the same neighborhood property to generate the super-resolution image for the rest of the regions. In effect what it does is as follows. If the wide angle view corresponds to a field of view of α^o, and the most zoomed view corresponds to a field of view of β^o (where $\alpha > \beta$), we generate a picture of the α^o field of view at a spatial resolution comparable to β^o field of view. The details of the method are now presented.

2. Low resolution image formation model

The zooming based super-resolution problem is cast in a restoration framework. There are p observed images $\{Y_i\}_{i=1}^{p}$ each captured with different zoom settings and of size $M_1 \times M_2$ each. Figure 1 illustrates the block schematic of how the low resolution observations at different zoom settings are related to the high resolution image. Here we consider that the most zoomed observed image of the scene Y_p has the highest resolution. We assume that the zoom factors between successive observations are known. It is also assummed that there is no change in the viewing direction, *i.e*, the optical axis remains the same during the zooming process. We further assume that there is no depth related perspective distortion due to the thick lens behavior of the camera. Since different zoom settings give

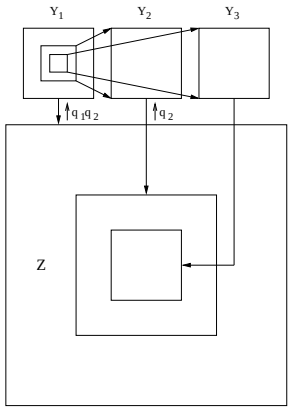

Figure 1: Illustration of observations at different zoom levels, Y_1 corresponds to the least zoomed and Y_3 to the most zoomed images. Here z is the high resolution image.

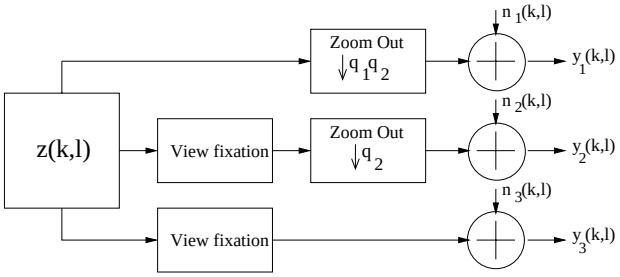

Figure 2: Low resolution image formation model. View fixation just crops a small part of the high resolution image z.

rise to different resolutions, the least zoomed scene corresponding to entire scene need to be upsampled to the size of $(q_1 q_2 \cdots q_{p-1}) \times (M_1 \times M_2) = (N_1 \times N_2)$, where $q_1, q_2, \cdots, q_{p-1}$ are the zoom factors between observed images of the scene $Y_1Y_2, Y_2Y_3, \cdots, Y_{(p-1)}Y_p$. Given Y_p, the remaining $(p-1)$ observed images are then modeled as decimated and noisy versions of this single high resolution image of the appropriate region in the scene. With this the most zoomed observed image will have no decimation. If $\underline{\mathbf{y}}_m$ is the $M_1 M_2 \times 1$ lexicographically ordered vector containing pixels from differently zoomed images Y_m, the observed images can be modeled as (refer to figure 2)

$$\underline{\mathbf{y}}_m = D_m \underline{\mathbf{z}} + \underline{\mathbf{n}}_m, \qquad m = 1, \cdots, p \tag{1}$$

where D is the decimation matrix, size of which depends on zoom factor. For a zoom factor of q, decimation matrix D consists of $\frac{1}{q^2}$ and has the form

$$D = \frac{1}{q^2}\begin{bmatrix} 11\ldots1 & & 0 \\ & 11\ldots1 & \\ & \ddots & \\ 0 & & 11\ldots1 \end{bmatrix} \quad (2)$$

p is the number of observations, $\underline{\mathbf{n}}_m$ is the $M_1M_2 \times 1$ noise vector. We assume noise to be zero mean i.i.d, and hence the multivariate noise probability density is given by

$$P(\underline{\mathbf{n}}_m) = \frac{1}{(2\pi{\sigma_\eta}^2)^{\frac{M_1M_2}{2}}}\exp\left\{-\frac{1}{2\sigma_\eta^2}\underline{\mathbf{n}}_m^T\underline{\mathbf{n}}_m\right\} \quad (3)$$

where σ_n^2 denotes the variance of noise process. Our problem now reduces to estimating $\underline{\mathbf{z}}$ given $\underline{\mathbf{y}}_m$'s, which is clearly an ill-posed, inverse problem.

3. MRF approach to super-resolve a scene

3.1. Stochastic models of Fields:

Markov random fields (MRFs) have emerged as a popular stochastic model for images due to its ability to capture local dependencies and its equivalence to the Gibbs random field (GRF) [4]. Let X be a random field over an arbitrary $N \times N$ lattice of sites $L = \{(i,j)|1 \le i,j \le N\}$. From the Hammersley-Clifford theorem [2] which proves the equivalence of an MRF and a GRF, we have $P(X = x) = \frac{1}{Z}e^{-U(x)}$ where x is an instance of X, Z is the partition function given by $\sum_x e^{-U(x)}$ and $U(x)$ is the energy function given by $U(x) = \sum_{c\in\mathcal{C}} V_c(x)$. $V_c(x)$ denotes the potential function of clique c and $\mathcal{C}$ is the set of all cliques.

The lexicographically ordered high resolution image $\underline{\mathbf{z}}$ satisfying the Gibbs density function is now written as

$$P(\underline{\mathbf{z}}) = \frac{1}{Z}\exp\left\{-\sum_{c\in\mathcal{C}} \mathrm{V}_c(\underline{\mathbf{z}})\right\} \quad (4)$$

In order to employ a simple and fast minimization technique like gradient descent, it is desirable to have a convex energy function. To this end we consider pair wise cliques on a first order neighborhood and impose a quadratic cost which is a function of finite difference approximations of the first order derivative at each pixel location, i.e.,

$$\begin{aligned}\sum_{c\in\mathcal{C}} V_c(\underline{\mathbf{z}}) &= \frac{1}{\lambda}\sum_{k=1}^{N_1}\sum_{l=1}^{N_2}[(z_{k,l} - z_{k,l-1})^2 \\ &+ (z_{k,l} - z_{k-1,l})^2]\end{aligned}$$

where λ represents the penalty for departure from smoothness in $\underline{\mathbf{z}}$

It is well known that in images, pixels with significacant change in intensities carry important information. In order to incorporate provisions for detecting such discontinuities, Geman and Geman [7] introduced the concept of line fields located on a dual lattice. The horizontal line field element $l_{i,j}$ connecting site (i,j) to $(i-1,j)$ aids in detecting a horizontal edge while the vertical line field element $v_{i,j}$ connecting site (i,j) to $(i,j-1)$ helps in detecting a vertical edge. We have chosen $l_{i,j}$ and $v_{i,j}$ to be binary variables over the line fields L and V. The on-state of the line-process variable indicates that a discontinuity, in the form of a high gradient, is detected between neighboring points, e.g., $l_{i,j} = 1$ if $|x_{i,j} - x_{i-1,j}| >$ *Threshold*, else $l_{i,j} = 0$. Similarly $v_{i,j} = 1$ if $|x_{i,j} - x_{i,j-1}| >$ *Threshold*, else $v_{i,j} = 0$. Each turn-on of a line-process variable is penalized by a quantity γ so as to prevent spurious discontinuities. Thus the energy function for the random process X with discontinuity fields L and V is written as [12]

$$\begin{aligned}U(x,l,v) &= \sum_{c\in\mathcal{C}} V_c(x,l,v) \\ &= \sum_{i,j}\mu[(x_{i,j} - x_{i,j-1})^2(1 - v_{i,j}) \\ &+ (x_{i,j+1} - x_{i,j})^2(1 - v_{i,j+1}) \\ &+ (x_{i,j} - x_{i-1,j})^2(1 - l_{i,j}) \\ &+ (x_{i+1,j} - x_{i,j})^2_(1 - l_{i+1,j})] \\ &+ \gamma[l_{i,j} + l_{i+1,j} + v_{i,j} + v_{i,j+1}] \\ &= \sum_{i,j}[\mu e_{xs} + \gamma e_{xp}]\end{aligned}$$

where e_{xs} and e_{xp}, respectively, are the smoothness term and the penalty term necessary to prevent occurance of spurious discontinuities. Here μ represents the penalty term for departure from the smoothness. We use this particular energy function in our studies in order to preserve discontinuities in the restored image. Any other form of energy function can also be used without changing the solution modality proposed here.

3.2. Maximum *a posteriori* (MAP) Solution

In order to use the maximum *a posteriori* (MAP) estimation technique to obtain the high resolution image $\underline{\mathbf{z}}$ given the ensemble of images at different resolutions we need to obtain

$$\hat{\mathbf{z}} = \arg\max_{\underline{\mathbf{z}}} P(\underline{\mathbf{z}} \mid \underline{\mathbf{y}}_1, \underline{\mathbf{y}}_2, \cdots, \underline{\mathbf{y}}_p) \quad (5)$$

Using Bayes' rule and taking the log of posterior probability

$$\hat{\mathbf{z}} = \arg\max_{\underline{\mathbf{z}}}\left[\log P(\underline{\mathbf{y}}_1, \underline{\mathbf{y}}_2, \cdots, \underline{\mathbf{y}}_p \mid \underline{\mathbf{z}}) + \log P(\underline{\mathbf{z}})\right] \quad (6)$$

$$= \arg\max_{\underline{\mathbf{z}}}\left[\sum_{m=1}^{p}\log P(\underline{\mathbf{y}}_m \mid \underline{\mathbf{z}}) + \log P(\underline{\mathbf{z}})\right], \quad (7)$$

since $\underline{\mathbf{n}}_m$ are independent. Now using equations (1) and (3), we get

$$P(\underline{\mathbf{y}}_m \mid \underline{\mathbf{z}}) = \left[\frac{1}{(2\pi {\sigma_\eta}^2)^{\frac{M_1 M_2}{2}}} \exp\left\{ -\frac{\|\underline{\mathbf{y}}_m - D_m \underline{\mathbf{z}}\|^2}{2{\sigma_\eta}^2} \right\} \right] \quad (8)$$

We model the scene to be recovered as an MRF. This is justified because the change in intensities in a scene is usually gradual and hence there is a local dependency. Thus using equation (4) and substituting in equation (7) the final cost function is obtained as

$$\hat{\mathbf{z}} = \arg \min_{\underline{\mathbf{z}}} \left[\sum_{m=1}^{p} \frac{\|\underline{\mathbf{y}}_m - D_m \underline{\mathbf{z}}\|^2}{2{\sigma_\eta}^2} + \sum_{c \in \mathcal{C}} V_c(\underline{\mathbf{z}}) \right]. \quad (9)$$

The above cost function is convex and is minimized using the gradient descent technique. The initial estimate $\underline{\mathbf{z}}^{(0)}$ is obtained as follows. Pixels in zero-order hold of the least zoomed observed image corresponding to the entire scene is replaced successively at appropriate portions with zero-order hold of the other observed images with increasing zoom factors. Finally the most zoomed observed image with highest resolution is copied with no interpolation.

In order to preserve discontinuities we modify the cost for prior probability term as discussed in Section 3 and use simulated annealing to minimize the cost. The cost function to be minimized then becomes

$$\hat{\mathbf{z}} = \arg \min_{\underline{\mathbf{z}}} \left[\sum_{m=1}^{p} \frac{\|\underline{\mathbf{y}}_m - D_m \underline{\mathbf{z}}\|^2}{2\sigma_\eta^2} + V(\underline{\mathbf{z}}) \right], \quad (10)$$

where

$$V(\underline{\mathbf{z}}) = \sum_{i,j} \left[\mu e_{zs} + \gamma e_{zp} \right]. \quad (11)$$

On inclusion of line fields in cost function, the gradient descent technique is liable to get trapped in local minima. Hence we minimize the cost by using simulated annealing (SA) which guarantees the attainment of global minima. However, in order to speed up the computation, the estimate obtained using the gradient descent method is used as the initial guess.

4. Experimental Results

In this section, we present the efficacy of the proposed techniques to recover the super-resolved image from scenes at different zooms through some examples on real data. In our experiments we used a zoom factor of 2 between images $Y_1 Y_2$ and 4 between $Y_1 Y_3$. Thus the zoom factor between $Y_2 Y_3$ becomes 2. Figures 3 (a-c) show input (observed) images of a house Y_1 , Y_2 , Y_3 each of size 72×96 with a zoom factor of 2 between images (a) and (b) and also a factor of 2 between (b) and (c). The automatic gain control (AGC) in the camera automatically set the camera gain in accordance with the amount of light in the pictured area and the level of zooming. Since we are capturing regions with different zoom setting, the AGC of the camera yields different average brightness for differently zoomed observation. Hence in order to compensate for AGC effect we used mean correction so as to maintain the average brightness of the captured images approximately the same. This is done for Y_2 image by subtracting its mean from each pixel and adding the mean due to its corresponding portion in Y_1 (refer to figure 1). Similarly for the Y_3 image we subtract its mean and add mean of its portion in Y_1. We used mean corrected images in all our experiments. Figure 4 shows the zoomed house image of size 288×384 obtained by bilinear interpolation of the least zoomed observed image Y_1 of size 72×96 with 144×192 sized bilinear interpolation obtained with Y_2 replacing that part of interpolated least zoomed observed image and 72×96 sized most zoomed observed image replacing those corresponding pixels in interpolated Y_1. The super-resolved image is shown in figure 5. Comparison of the figures show less blockiness in the super-resolved image. The seam is clearly visible in figure 4. Also the branches in the plants are more clearly distinguishable in the super-resolved image. The values of the parameters used in simulated annealing (SA) algorithm for recovering house image are $\mu = 0.0095$, $\gamma = 150$, *Threshold* $= 25$, the decrement factor for temperature in the annealing schedule $\delta = 0.999$, initial temperature $T_0 = 3.15$.

Next we consider the case where the observed images Figure 6 (a-c) have less intensity variations. Figures 7, and 8 show the corresponding results with the girl image. Again notice the seam in figure 7. The super-resolution image corresponding to entire scene Y_1 consists of super-resolved image due to Y_2 which has super-resolved image Y_3. In the case of girl image Y_1, Y_2 and Y_3 correspond to figure 6 (a), (b), and (c). In this experiment we found discontinuities were better preserved by considering three different *Threshold* values for the three super-resolution regions, and the parameters used were $\mu = 0.08$, $\gamma = 10$, *Threshold1* $= 70$ for the super-resolution region of Y_1 only, *Threshold2* $= 10$ for the super-resolution region of Y_2 only, and *Threshold3* $= 5$, for the super-resolved region Y_3. The justification for selecting different thresholds for different regions is that one has gradually less information about the peripheral regions, and hence the restoration tends to be more smooth at these regions. The choice of lower values of threshold tries to prevent oversmoothing. The value for δ was chosen as 0.99. Initial temperature T_o was set to 3.15, same as before. The initial estimate for the high resolution image in these experiments was chosen to be the output obtained by using the gradient descent technique to reduce the computational burden.

Comparison of these results demonstrate that the pro-

(a) (b) (c)

Figure 3: Observed images of a house captured with three different zoom settings.

Figure 4: Zoomed house image formed by successive bilinear expansion as illustrated in figure 1.

Figure 5: Super-resolved house image.

(a) (b) (c)

Figure 6: Observed images of a girl captured with three different zoom settings.

posed method does yield a better result. However, the restotaion tends to be a bit too smooth near the periphery. This is quite expected as we have used just three observations and the peripheral region has been upsampled by a factor of 4×4. By use of more number of observations, we are likely to perform better. The effect of over smmothness is quite visible in the house image where the high frequency region corresponding to the plants appear to be very smooth. But the picture of the girl when super-resolved does not show this effect that prominently. The result does appear visually quite pleasant. The simulated annealing optimization method used here is quite slow and in order to decrease the computation time we implemented the mean field annealing optimization (MFA). The recovered super-resolved girl image using MFA is shown in figure 9. This compares quite favorably with the results given in figure 8.

5. Conclusions

We have presented a technique to recover the super-resolution intensity field from a sequence of zoomed observations. The resolution of the entire scene is obtained at the resolution of the most zoomed observed image which consists of only a portion of the actual scene. The super-resolved image is modeled as an MRF and a MAP estimate is used to derive the cost function to be minimized. The basic idea is that the entire observation conforms to the same MRF, but viewed at the different resolution pyramid. A simulated annealing optimization algorithm is used to minimize the cost. In order to increase the speed of convergence we try the mean field annealing scheme which uses deterministic annealing for a faster convergence and attains a near global minima. At present we have not considered the implementation in realtime. We have assumed here that zoom factors are known and considered only integer zoom factors. A more realistic situation is one in which it is unknown. Our future work involves simultaneously estimating the unknown zoom factors between observed images (scenes) while recovering the super-resolved image.

References

[1] S. Baker and T. Kanade. Super-resolution: Limits and beyond. In Subhasis Chaudhuri, editor, *Super-Resolution Imaging*, pages 244–276. Kluwer Academic Publisher,Boston, 2001.

[2] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of Royal Statistical Society, Series B*, 36:192–236, 1974.

[3] M.-C. Chiang and T. E. Boult. Local blur estimation and super-resolution. In *Proc. IEEE Conf. Computer Visin and Pattern Recognition*, pages 821–826, Puerto Rico, USA, 1997.

Figure 7: Zoomed girl image formed by using successive bilinear expansion.

Figure 8: Super-Resolved girl image.

[4] R. C. Dubes and A. K. Jain. Random field models in image analysis. *Journal of Applied Statistics*, 16(2):131–164, 1989.

[5] M. Elad and A. Feuer. Restoration of a single super-resolution image from several blurred, noisy and undersampled measured images. *IEEE Trans. on Image Processing*, 6(12):1646–1658, December 1997.

[6] M. Elad and A. Feuer. Super-resolution restoration of an image sequence : Adaptive Filtering approach. *IEEE Trans. on Image Processing*, 8(3):387–395, March 1999.

[7] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of image. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[8] Hassan Shekarforoush, Marc Berthod, Josiane Zerubia and M. Werman. Sub-pixel Bayesian estimation of albedo and height. *International Journal of Computer Vision*, 19(3):289 – 300, 1996.

Figure 9: Super-Resolved girl image using mean field annealing optimization.

[9] M. Irani and S. Peleg. Improving resolution by image registration. *CVGIP:Graphical Models and Image Processing*, 53:231 – 239, March 1991.

[10] S. P. Kim, N. K. Bose, and H. M. Valenzuela. Recursive reconstruction of high resolution image from noisy undersampled multiframes. *IEEE Trans. on Accoustics, Speech and Signal Processing*, 18(6):1013–1027, June 1990.

[11] D. Rajan and S. Chaudhuri. A Generalized Interpolation Scheme for Image Scaling and Super-resolution. In *Workshop on Vision, Models and Visualization*, pages 301–308, Erlangen,Germany, Nov 1999.

[12] D. Rajan and S. Chaudhuri. An MRF based approach to Generation of Super-resolution images from Blurred Observations. In *International Workshop on Vision, Models and Visualization*, Saarbruecken,Germany, Nov 2000.

[13] R. R. Schultz and R. L. Stevenson. A Bayesian approach to image expansion for improved definition. *IEEE Trans. on Image Processing*, 3(3):233–242, May 1994.

[14] R. R. Schultz and R. L. Stevenson. Extraction of high-resolution frames from video sequences. *IEEE Trans. on Image Processing*, 5:996 – 1011, June 1996.

[15] C. Srinivas and M. D. Srinath. A stochastic model based approach for simultaneous restoration of multiple misregistered images. *SPIE*, 1360:1416 – 1427, 1990.

[16] Subhasis Chaudhuri. *(Ed.), Super-Resolution Imaging*. Kluwer Academic Publisher, Boston, 2001.

[17] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan. High resolution image reconstruction from lower-resolution image sequences and space-varying image restoration. In *Proc. ICAASP*, pages 169–172, San Francisco,USA, 1992.

[18] R. Y. Tsai and T. S. Huang. Multiframe image restoration and registration. In *Advances in Computer Vision and Image Processsing*, pages 317–339. JAI Press Inc., 1984.

[19] H. Ur and D. Gross. Improved resolution from sub-pixel shifted pictures. *CVGIP:Graphical Models and Image Processing*, 54:181 – 186, March 1992.

[20] T. William T. Freeman and E. C.Pasztor. Example- based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, March/april 2002.