

2009

Super resolution and dynamic range enhancement of image sequences

Lutfi Murat Gevrekci

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations



Part of the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Gevrekci, Lutfi Murat, "Super resolution and dynamic range enhancement of image sequences" (2009).
LSU Doctoral Dissertations. 3457.

https://digitalcommons.lsu.edu/gradschool_dissertations/3457

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.

SUPER RESOLUTION AND DYNAMIC RANGE ENHANCEMENT OF IMAGE SEQUENCES

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

Department of Electrical and Computer Engineering

by

Lutfi Murat Gevrekci
B.S., Bilkent University, 2004
M.S., Louisiana State University, 2007
May 2009

To my loving parents: Munire and Huseyin

ACKNOWLEDGMENTS

I would like to thank Dr. Bahadir K. Gunturk for his support, inspiration, and encouragement throughout my Ph.D. I would like to thank Microsoft real time collaboration department for giving me the chance to apply my video quality assessment knowledge on video over IP. I also thank Label Vision Systems for introducing me to the most challenging computer vision tasks.

Special thanks to Dr. Jianhua Chen and Dr. Suresh Rai for invaluable knowledge they provided in machine learning and neural networks. I also thank Dr. Jerry Trahan and Dr. Guoping Zhang for serving in my committee. I also would like to thank George Trian Amariuca for all intellectual discussions.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	v
LIST OF FIGURES	vi
ABSTRACT	xi
CHAPTER	
I	BACKGROUND 1
	A. Imaging Model 1
	B. Spatial Resolution Enhancement 6
	C. Dynamic Range Enhancement 15
	D. Image Registration 19
II	ILLUMINATION ROBUST INTEREST POINT DETECTION . 24
	A. Introduction 24
	B. Harris Corner Detector 26
	C. Contrast Invariant Feature Transform 27
	D. Experiments 31
	E. Conclusions 33
III	SUPER RESOLUTION UNDER PHOTOMETRIC DIVER- SITY OF IMAGES 38
	A. Introduction 38
	B. Photometric Modeling 40
	C. SR under Photometric Diversity 45
	D. Geometric and Photometric Registration 54
	E. Experiments and Results 56
	F. Conclusions 58
IV	RESTORATION OF BAYER-SAMPLED IMAGE SEQUENCES 62
	A. Introduction 62
	B. Imaging Model 65
	C. Multi-Frame Demosaicking 66
	D. Achieving Subpixel Resolution 73
	E. Experiments 75
	F. Conclusions 80
V	CONCLUSIONS 86
REFERENCES	89
APPENDIX A	ITERATIVE IMAGE ENHANCEMENT 96
APPENDIX B	SUBBAND IMAGE DECOMPOSITION 98
VITA	100

ABSTRACT

Camera producers try to increase the spatial resolution of a camera by reducing size of sites on sensor array. However, shot noise causes the signal to noise ratio drop as sensor sites get smaller. This fact motivates resolution enhancement to be performed through software. Super resolution (SR) image reconstruction aims to combine degraded images of a scene in order to form an image which has higher resolution than all observations. There is a demand for high resolution images in biomedical imaging, surveillance, aerial/satellite imaging and high-definition TV (HDTV) technology. Although extensive research has been conducted in SR, attention has not been given to increase the resolution of images under illumination changes. In this study, a unique framework is proposed to increase the spatial resolution and dynamic range of a video sequence using Bayesian and Projection onto Convex Sets (POCS) methods. Incorporating camera response function estimation into image reconstruction allows dynamic range enhancement along with spatial resolution improvement. Photometrically varying input images complicate process of projecting observations onto common grid by violating brightness constancy. A contrast invariant feature transform is proposed in this thesis to register input images with high illumination variation. Proposed algorithm increases the repeatability rate of detected features among frames of a video. Repeatability rate is increased by computing the autocorrelation matrix using the gradients of contrast stretched input images. Presented contrast invariant feature detection improves repeatability rate of Harris corner detector around %25 on average. Joint multi-frame demosaicking and resolution enhancement is also investigated in this thesis. Color constancy constraint set is devised and incorporated into POCS framework for increasing resolution of color-filter array sampled images. Proposed method provides fewer demosaicking artifacts compared to existing POCS method and a higher visual quality in final image.

CHAPTER I

BACKGROUND

In this section we present essential background information for understanding resolution and dynamic range enhancement. First we investigate the image formation model and the artifacts introduced through imaging pipeline. Section B and Section C provides different resolution and dynamic range enhancement techniques, respectively, along with existing literature on the subjects. We also provide image registration techniques in Section D, which is fundamental for multi-frame image enhancement algorithms since images taken from different point of views should be aligned onto a common grid.

A. Imaging Model

Image formation is the process where scene radiance is converted to intensity through camera. Image formation consists of several interdependent steps, each of which can introduce several artifacts. Although isolating image formation steps in terms of artifacts is a cumbersome task, we broadly divide image acquisition pipeline into preprocessing and postprocessing phases.

Lens, aperture rate, shutter speed, sensor array, ISO (sensor array gain), and camera response function are the camera components and adjustment units (parameters) included in preprocessing step. These components are depicted in Figure 1 along with the artifacts introduced. Note that we assumed a Bayer sensor array is used in the system.

The journey of pixels start by scene radiance passing through aperture then through camera lens. Aperture behaves as the eye pupil and determines the amount of light entering camera system. Increase in aperture size might lead not only to decrease in depth of field but also vignetting in some camera systems. Lens is one of the most important components in imaging pipeline in terms of visual quality. Improper lenses introduce optical blur which leads to sharpness (resolution) loss. Color fringing is

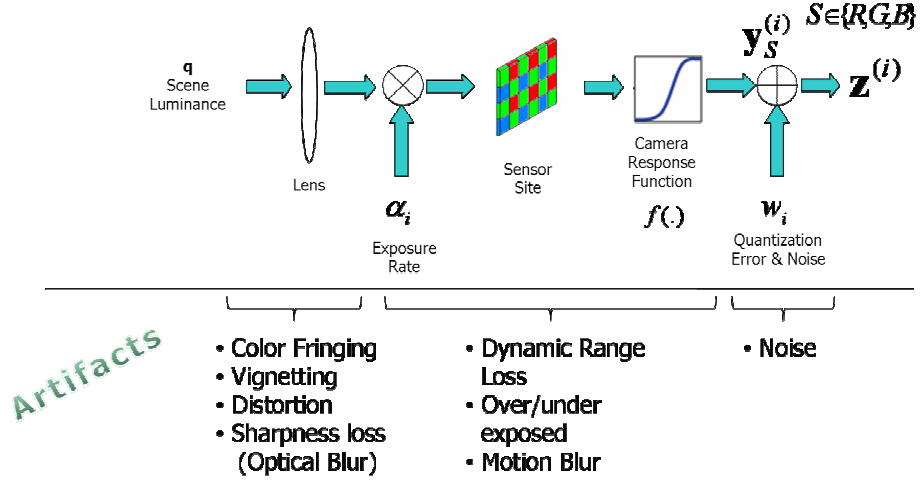


Fig. 1.: Preprocessing Phase of Image Formation Model

another possible lens flaw. Incoming light has several spectral components. A proper lens system should project light onto on optical axis without refracting the spectral components to different locations. Lack of this behavior creates color fringing artifacts in final image. Optical aberration occurs when a lens system radially distorts the ground truth geometry. Most common types of optical aberrations are barrel and pincushion distortions. Barrel distortion is a common defect observed in fish eye lenses. Radiance that passes the lens system is finally projected onto sensor array and collected as photons in each sensor site. Integration time of photons at a sensor site is determined by the shutter speed. Inadequate shutter speed may lead to motion blur in case of high speed movement of an object in the scene. Shutter speed is known as exposure duration as well. Exposure duration is represented along with ISO as a multiplicative term (exposure rate) in Figure 1 for visualization purposes. Exposure duration can be manually adjusted by the user or automatically set by the camera system. Improper exposure adjustment can lead to over/under exposed images. High exposure factor can boom up noise since noisy low range values will be amplified by this process. Shutter speed has an influence on effective frame rate during video mode since it determines integration time of each frame. Photons collected at a sensor site during integration time are multiplied by sensitivity (ISO) and converted to electrical

charge through camera response function a.k.a opto-electronic conversion function. Camera response function has generally a non-linear nature to form saturated colors during photometric conversion[4]. Camera response function not only manipulates the colors but also define the dynamic range of the final image. Dynamic range is basically the existing contrast ratio in an image. Dynamic range of an observation depends on the camera response function, shutter speed, ISO and sensor site dimensions to some extent. Cameras with larger sensor site sizes yield higher dynamic range allowing to collect more number of photons.

At the end of preprocessing phase of image formation we obtain a quantity called exposure value, which can be considered as amount of light received by camera. Exposure value is determined according to aperture, shutter speed, sensitivity, and camera response function. Additive imaging noise is introduced by exposure rate, ISO and possibly by temperature.

Postprocessing phase takes raw Bayer image as input and creates three channel color image to be stored on camera as illustrated in Figure 2. This assumption is valid unless a beam splitter is used to decompose incoming light into three spectral components. Note that there are *3CCD* sensor arrays, and three imaging pipelines in such systems. Instead most cameras produce each spectral component from a single mosaicked filter array, which is denoted as color filter array (CFA) interpolation or commonly known as demosaicking. A famous color filter is shown in first stage of Figure 2, which is known as Bayer pattern.

Color filter array (CFA) mosaic stores one color component at each sensor site. CFA interpolation decreases the spatial resolution as one color value is present for each location. Missing color components are estimated from mosaic patten by interpolation. Interpolation leads to sharpness loss, and to decrease in resolution consecutively. Demosaicking step might introduce color artifacts on highly textured regions during interpolation. Cameras also perform white balancing on extracted color components

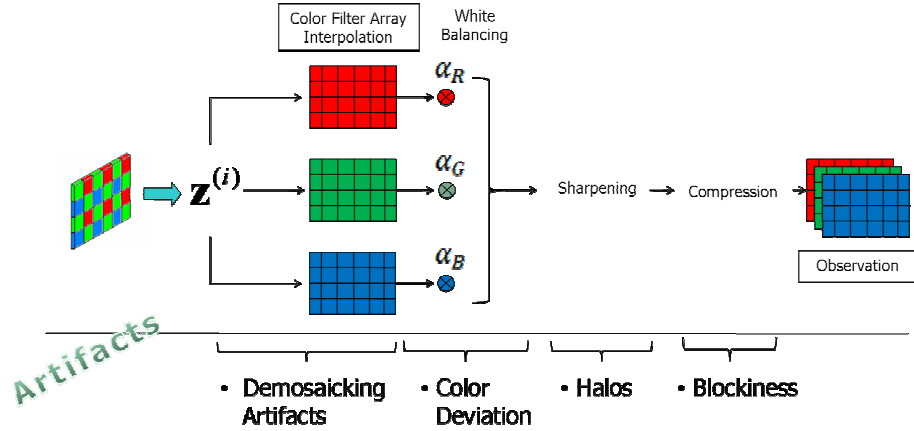


Fig. 2.: Postprocessing Phase of Image Formation Model

to remove unrealistic colors from final image. Color channels are weighted with different coefficients depending on the color temperature estimate to achieve color balance and preserve *true white*. Each camera also has internal sharpening to overcome the blur introduced by the point spreading function of the lens system. Internal over sharpening can introduce halos at edge locations, while the lack of sharpening may result in a blurry image. Final image can be saved with or without compression depending on the selection of user. There are cases where compression is applied to the final image as default i.e cell phone cameras. Block based compression is applied at the final step. This creates blocking artifacts at block border and lack of variance within each block depending on the percentage of discrete cosine transform (DCT) coefficients discarded. Three channel image is saved in memory after going through preprocessing and postprocessing image formation phases.

Image restoration aims to recover less degraded version of a scene from a single image with aforementioned blemishes. This approach is limited since accurate information may be impossible to retrieve from single image while multi-frame image fusion provides higher immunity to artifacts due to abundance of extra information. Diversity of information coming from multiple images allows to compensate for the degradations and artifacts. Extra information might originate from redundancy in

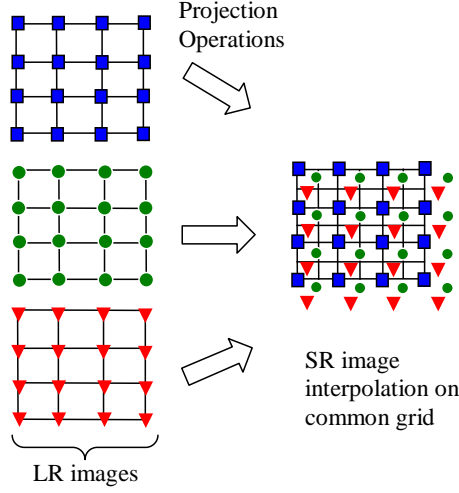


Fig. 3.: High resolution formation on common grid. Images on left side represent LR images. Projecting all samples onto common grid and interpolating on this grid yields resolution increase.

geometric transformation or different tonal range each observation possesses. Multi-frame techniques are used in literature for noise denoising, demosaicking, dynamic range enhancement. Panoramic imaging is also a multi-frame fusion method to compensate for the camera angle of view. Super resolution is the name of multi-frame image fusion technique to increase the inherent (spatial) resolution of an observation.

Previous work on super resolution is based on imaging model that incorporates only warping, blurring and spatial sampling effects. Blurring and sampling converts samples lying on a high resolution grid to a low resolution one, while warping changes the sampling location. Here samples are averaged by a point spreading function (PSF) and sampled at a certain rate. Aliasing occurs if sampling rate is insufficient. Compensating warping, blurring and spatial sampling in reconstruction allows us to obtain only spatial resolution enhancement, without any tonal (range) resolution improvement and with possible color artifacts. In this thesis, effect of non-linear camera response function, exposure rate changes and multi-frame demosaicking techniques are incorporated into image enhancement framework. This technique yields results with improved spatial and dynamic resolution, with less color artifacts.

B. Spatial Resolution Enhancement

In spatial resolution enhancement, imaging process is assumed to consist of geometric transformation, blurring, and additive noise. An oversimplified case of spatial SR enhancement can be considered when one of the observations is chosen as reference frame, onto which each observation is geometrically projected, and images are non-uniformly interpolated on this common grid. This process is depicted in Figure 3. In this figure, LR images are shown on left side. During reconstruction, circle shaped LR image is chosen as reference on which multi-frame non-uniform interpolation is performed. In spite of its simplicity, this model is useful to demonstrate the importance of registration. Accurate registration is crucial for the success of SR as sub-pixel precision is required to increase resolution.

As explained in Section B, blurring may also occur either because of the optical system (out of focus, diffraction limit) or due to motion (inadequate sampling rate, fast movement in the scene). Point spreading function (PSF) of the sensor system causes the discrete valued ground truth image get spatially averaged, and PSF shape may locally vary depending on motion blur. Spatial sampling is another factor that decreases the resolution of warped and blurred image. Color filter array (CFA) interpolation causes further spatial sampling where a single CCD camera is used. Set of observations degraded with aforementioned artifacts is provided in Figure 4 (a) - (c). Figure 4 (d) has the super resolution (SR) image, formed by multi-frame image enhancement. Notice that in final SR image artifacts are removed while sharpness is increased.

First we will present literature on SR assuming color channel information is available without need for CFA interpolation. SR reconstruction from CFA sampled (raw) images will be provided in Chapter IV. For mathematical consistency we represent the LR image formation in matrix form. Denoting \mathbf{x} as SR image, k th LR image \mathbf{y}_k



Fig. 4.: Spatial resolution enhancement of degraded images. 37 images are used in this experiment, 3 of which are shown in (a),(b),(c). SR image is given in (d).

can be formed from SR image as following

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{n}_k, \quad (1.1)$$

where \mathbf{H}_k is the multiplication of downsampling, blurring and warping matrices $\mathbf{H}_k = \mathbf{D}\mathbf{B}_k\mathbf{M}_k$. Using the notation of [5], matrices are given in Table 1. In this table, L_1 and L_2 are the downsampling factors in horizontal and vertical directions. Number of SR pixels is $N = L_1 N_1 L_2 N_2$, and number of LR pixels is denoted as $M = N_1 N_2$. Spatial sampling, which is the main source of aliasing in observations, is represented by downsampling. Here downsampling (\mathbf{D}) is assumed to be same for all images, leading to a constant enhancement factor in reconstruction. In case of having a color filter array in imaging process, color filtering will also bring aliasing. Leaving color filtering to discuss later in Chapter IV, principal components of imaging process can be enumerated as geometric transformation, optical blur, and spatial sampling (downsampling). In following paragraphs we present the existing SR work in literature.

Table 1.: Matrix representation of imaging model parameters

High Resolution image	$\mathbf{x} = [x_1, x_2, \dots, x_N]^T$	$L_1 N_1 L_2 N_2 \times 1$
k th Observation	$\mathbf{y}_k = [y_{k,1}, y_{k,2}, \dots, y_{k,M}]^T$	$N_1 N_2 \times 1$
Warp Matrix for k th image	\mathbf{M}_k	$L_1 N_1 L_2 N_2 \times L_1 N_1 L_2 N_2$
Downsampling Matrix	\mathbf{D}	$(N_1 N_2)^2 \times L_1 N_1 L_2 N_2$
Blur Matrix for k th image	\mathbf{B}_k	$L_1 N_1 L_2 N_2 \times L_1 N_1 L_2 N_2$
Joint Matrix	\mathbf{H}_k	$(N_1 N_2)^2 \times L_1 N_1 L_2 N_2$

First multi-frame resolution enhancement algorithm is proposed by Tsai and Huang [6] to be performed in frequency domain. This approach relates the discrete fourier transform (DFT) coefficients of LR and continuous fourier transform (CFT) of SR image using shift invariance of fourier transform. Although this method is computationally efficient, it has some drawbacks. Frequency based SR can only accommodate translational motion and LSI blur. Other frequency based reconstruction methods are [7, 8, 9, 10]. This approach is then modified to incorporate different types of motion and regularization parameters.

SR image can be obtained by solving the imaging model given in Equation 1.1 by least squares. In case additive noise is zero mean and Gaussian, problem simplifies to maximum likelihood (ML) estimation. ML approach solves for SR without any regularization term. ML solution can be found by the following optimization

$$\mathbf{x}_{mle} = \arg \min_x \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2, \quad (1.2)$$

where $\mathbf{y} = [\mathbf{y}_0^T \mathbf{y}_1^T \dots \mathbf{y}_{N-1}^T]^T$, $\mathbf{H} = [\mathbf{H}_0^T \mathbf{H}_1^T \dots \mathbf{H}_{N-1}^T]$, and N is the number of observations. Least squares solution of this system is $\hat{\mathbf{x}}_{mle} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}$. Solving SR image by matrix manipulation is inefficient due to size of matrices. Alternatively, iterative techniques are used to find SR image. An efficient ML solution of SR is proposed by Irani and Peleg in [11]. Later Irani and Peleg extended this approach

to handle occlusion [12]. Tom and Katsaggelos [13] used ML to estimate SR image and noise variance together. ML solution has the drawback of noise amplification.

Regularization is proposed in stochastic reconstruction approaches to suppress noise that occurs in ML solution. Stochastic reconstruction method is mostly preferred due its easiness to adopt prior knowledge about solution. Bayesian approach, which is also denoted as maximum a-posteriori (MAP), is the most commonly used probabilistic method when the distribution of the solution is available. Prior knowledge in MAP acts as a regularization term leading to a stabilized result. Tikhonov regularization is one of the earliest form of prior knowledge [14, 15]. Hardie et al. modified MAP scheme to estimate SR and registration simultaneously [16]. Cheese-man used MAP to increase the resolution of satellite images [17]. Markov random field (MRF) is a common technique to incorporate prior knowledge into MAP solution. Gaussian MRF provides prior knowledge by containing n th order difference operator. Schultz and Stevenson [18] proposed second order derivative for SR reconstruction. Another constraint is Huber MRF which provides further edge preservation compared to Gaussian MRF, which is also given in [18]. Huber MRF provides the sharper results by giving heavier weight to tails compared to the Gaussian weight.

Also scene dependent priors can be used for specific computer vision tasks. Capel and Zisserman [19] proposed specific priors for face recognition. Similarly, Gunturk et al. [20] proposed a Gaussian prior for face recognition. MAP has the advantage of incorporating any prior knowledge including the noise statistics into the solution.

Bayesian estimation modifies ML approach by allowing the incorporation of prior knowledge. In Bayesian estimation, we seek SR image (\mathbf{x}) given the observations (\mathbf{y}). This leads us to find the following conditional distribution $P(\mathbf{x}|\mathbf{y})$ to solve for SR image. Using Bayes rule, this probability can be written in following form $P(\mathbf{x}|\mathbf{y})=P(\mathbf{y}|\mathbf{x}) P(\mathbf{x})/P(\mathbf{y})$. Assuming that all observations are equally likely to ap-

Table 2.: Priors used for image regularization

Median Prior	$P(\mathbf{x}) = \frac{1}{Z} \exp \left(-\frac{ \mathbf{x} - \mathbf{x}_{\text{med}} }{2\sigma_x^2} \right)$
MRF Prior	$P(\mathbf{x}) = \frac{1}{Z} \exp \left(-\sum_{c \in S} \varphi_c(\mathbf{x}) \right)$
Gaussian MRF Prior	$\varphi_c(\mathbf{x}) = (D^n(\mathbf{x}))^2$
Huber MRF Prior	$\varphi_c(\mathbf{x}) = \begin{cases} \mathbf{x}^2, & \mathbf{x} \leq \alpha \\ 2\alpha \mathbf{x} - \alpha^2, & \text{o.w} \end{cases}$

pear, MAP solution can be found by optimizing the cost function below

$$\begin{aligned} \mathbf{x}_{\text{map}} &= \arg \max_x \log P(\mathbf{x}) + \log P(\mathbf{y}|\mathbf{x}) \\ &= \arg \max_x \log P(\mathbf{x}) - \frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 \end{aligned} \quad (1.3)$$

MAP approach differs according to the prior model used.

$$P(\mathbf{x}) = \frac{1}{Z} \exp \left(-\mathbf{x}^T L^T L \mathbf{x} \right), \quad (1.4)$$

where Z is a normalizing constant. In this case, MAP can be found by optimizing the following

$$\mathbf{x}_{\text{map}} = \arg \min_x \gamma^2 \|\mathbf{L}\mathbf{x}\|_2^2 + \sum_{k=1}^N \|\mathbf{y}_k - \mathbf{H}_k \mathbf{x}\|_2^2, \quad (1.5)$$

where \mathbf{L} can be the discrete approximation of a derivative operator, Laplacian or identity matrix. γ in Equation A.1 is the regularization parameter. Here regularization term provides a smoother solution for SR image. Regularization parameter (γ) should be chosen carefully as smoothing operations acting on edges to remove the sharpness. Robust regularization parameter (γ) estimation is proposed in [21] using L-curve method.

PSF and regularization parameters are estimated together in [22] using generalized cross-validation. Importance of regularization becomes evident in case LR images are limited or the accuracy of the assumed imaging model is low. Other prior models are given in Table II, where $D^n(\cdot)$ represents the finite difference operator, α is the

cutoff value for Huber prior, and $\varphi_c(\cdot)$ is the potential function that is effective within spatial range of S .

Minimization of cost function in Equation A.1 leads to the following solution which is known as Tikhonov regularization $\mathbf{x}_{map} = (\mathbf{H}^T \mathbf{H} + \gamma^2 \mathbf{L}^T \mathbf{L})^{-1} \mathbf{H}^T \mathbf{y}$. Due to the size of the matrices involved we avoid taking matrix inverse. Instead we apply iterative methods to solve for SR image. Zomet and Peleg [23] gives a detailed description of conjugate gradient (CG) solution for SR image. SD and CG solution of SR image is given in Appendix A for derivative prior.

Another alternative SR reconstruction is proposed by Farsiu et al. [24], by representing the general cost function in L_1 norm instead of L_2 and using total variation (TV) as regularization. Previously the regularization term was proposed to be $\|\mathbf{L}\mathbf{x}\|_2$ while in TV the regularization term becomes $\|\nabla\mathbf{x}\|_1$. TV based regularization is claimed to preserve edges far superior than other methods [24]. Using bilateral filter to implement TV, regularization function becomes

$$\Upsilon_{BTV}(\mathbf{x}) = \sum_{l=-P}^P \sum_{m=-P}^P \alpha^{|m|+|l|} \|\mathbf{x} - S_x^l S_y^m \mathbf{x}\|_1, \quad (1.6)$$

where α is the regularization constant in the range of $[0,1]$. S_x and S_y are the shift operators in x and y directions, respectively. Incorporating the regularization into the overall cost function becomes

$$\mathbf{x} = \arg \min \left[\sum_{k=1}^N \|\mathbf{H}_k \mathbf{x} - \mathbf{y}_k\|_1 + \lambda \Upsilon_{BTV}(\mathbf{x}) \right] \quad (1.7)$$

Again this equation can be solved by iterative methods such as SD, or CG.

Projection onto convex sets (POCS) is another approach that solves for SR image which is consistent with predefined constraint sets. Different constraint sets are proposed to produce a result close to the ground truth image. Final image lies at the intersection of convex constraint sets. This can be achieved by applying the projection operations consecutively in an iterative fashion. Although final image is consistent

with applied constraint sets, final result is dependent on the order of the projections being applied. Applying m number of constraint sets on input image is formulated as following.

$$\mathbf{x}^{(k+1)} = P_m P_{m-1} \dots P_2 P_1 \mathbf{x}^{(k)}, \quad (1.8)$$

where \mathbf{x}^0 is the initial SR estimate, P_i is the projection operator according to the i th constraint set C_i , and k denotes the iteration number.

Data consistency constraint set is most commonly used to provide consistency with observations. This can be achieved by producing simulated observations from SR image thorough imaging model. Our aim is to keep the residual difference between simulated observations and actual observations small as possible. This constraint set is defined as following

$$C_i[y_i(l_1, l_2)] = \{x(n_1, n_2) : |r^{(x)}(l_1, l_2)| \leq T_i(l_1, l_2)\}, \quad (1.9)$$

where the residual is $r^{(x)}(l_1, l_2) = y_i(l_1, l_2) - \sum_{n_1, n_2} h_i(n_1, n_2; l_1, l_2)x(n_1, n_2)$. Threshold $T_i(l_1, l_2)$ is supposed to be related to the power of noise $n_i(n_1, n_2)$. Here, (n_1, n_2) is the discrete coordinate of SR image, whereas (l_1, l_2) is the coordinates of LR image. Similarly $y_i(l_1, l_2)$ represents the observation value and $x(n_1, n_2)$ represents the value of high resolution image at specified coordinates of (l_1, l_2) and (n_1, n_2) , respectively. Convolution kernel that produces simulated LR image out of SR image is represented as $h_i(\cdot)$. SR image can be updated in following fashion

$$x^{(k+1)}(n_1, n_2) = x^{(k)}(n_1, n_2) + \begin{cases} \frac{(r^x(\cdot) - T_i(\cdot))h_i(n_1, n_2; \cdot)}{\sum_{o_1} \sum_{o_2} h_i^2(o_1, o_2; \cdot)}, & r^x(\cdot) > T_i(\cdot) \\ 0, & T_i(\cdot) > r^x(\cdot) > -T_i(\cdot) \\ \frac{(r^x(\cdot) + T_i(\cdot))h_i(n_1, n_2; \cdot)}{\sum_{o_1} \sum_{o_2} h_i^2(o_1, o_2; \cdot)}, & r^x(\cdot) < -T_i(\cdot) \end{cases} \quad (1.10)$$

where “ \cdot ” represents the coordinates (l_1, l_2) and k is the iteration number. Although POCS is a deterministic approach, prior knowledge can still be incorporated in terms of additional constraint sets.

Originally POCS is proposed by Stark and Oskoui [25]. Tekalp et al. incorporated noise statistics into POCS method [26]. Change in aperture size of camera has been considered by Patti et al. to remove spatially varying blur using POCS [27]. Validity map is incorporated to compensate possible registration errors, and segmentation map is used to perform object based POCS reconstruction in [28]. Patti and Altunbasak updated the constraint sets according to the image structure and used higher order interpolants in POCS [29]. Although POCS is an effective reconstruction method, slow convergence and non-uniqueness of the solution are the drawbacks of this approach. Comparative results of MAP and POCS are provided in Figure 5. Results are obtained after five iterations, using 29 images. Resizing factor is chosen as two. You can notice that although POCS provides sharp results, ringing artifacts might occur around edges. MAP provides smoother results due to the regularization.

Hybrid ML/MAP/POCS is another reconstruction approach. This hybrid approach combines MAP and POCS, such that both prior knowledge and constraint sets are used to find SR solution. Schultz and Stevenson [18] made the earliest attempt in the hybrid approach. Elad and Feuer [30] proposed a similar hybrid approach. Hybrid approach both ensures uniqueness and takes advantage of the prior knowledge.

There are various other fields of research related to super-resolution imaging. Blur identification is one of the major research topics that aim to estimate the point spreading function (PSF) of the system. Generalized cross validation (GCV) is a common technique to estimate the PSF. Reeves and Mersereau demonstrates that GCV outperforms maximum likelihood in PSF selection [31]. Quality assessment of super resolution images in the absence is a complicated task due to absence of ground truth value. Performance estimation of super resolution is an active research area. Eeekeren et. al used triangle orientation discrimination as a performance measure [32].



(a)



(b)



(c)



(d)



(e)



(f)

Fig. 5.: POCS v.s MAP . (a)-(b) Low resolution images. (c)-(d) MAP results. (e)-(f) POCS results.

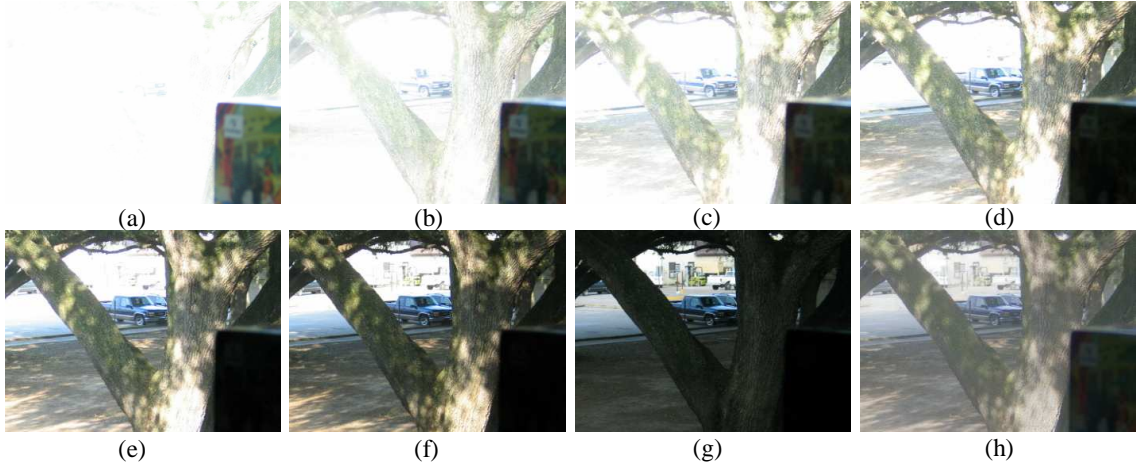


Fig. 6.: High dynamic image formation. (a)-(g) Low resolution images with varying exposure rates. (h) High dynamic image spanning the whole photometric range.

C. Dynamic Range Enhancement

Dynamic range enhancement has been performed in the name of HDR imaging in literature. *Dynamic range* is defined as the ratio of the highest and lowest value of illumination in the scene [33]. Due to the change in camera parameters (exposure time, aperture size, etc.) and tonal compression in camera systems, only a limited portion of the dynamic range can be captured in a single frame. HDR imaging aims to cover the existing tonal range in a scene by combining the images with different exposure times. An image set consisting of over and under exposed images are given in Figure 6 (a) through (g), and HDR result of this data set is given in Figure 6-(h).

HDR imaging has been studied extensively in literature. Dynamic range improvement necessitate estimation of camera response function (CRF). Debevec and Malik estimated CRF by minimizing the cost function consisting of radiance differences given the exposure durations [34]. Tsing et al. modeled imaging process using statistical calibration, and estimated both CRF and white balancing parameters jointly in an iterative manner [35]. Mitsunga and Nayar fit a polynomial function to CRF iteratively starting from a rough estimate of the exposure ratios of images [4]. Mann and Candocia approached tonal and geometric registration jointly [36, 37] using a parametric CRF. All HDR algorithms aim to find the radiance map of the scene us-

ing CRF. Visualization of radiance map is another issue as current display devices can be incapable of representing the HDR result. Many algorithms have been proposed to represent HDR result by range compression. Mainstream radiance map recovery and display approaches are given in survey of Battiato et al. [33]. We will present fundamental HDR approaches given in this survey. In essence, HDR algorithms aim to obtain camera independent scene radiance from the observations that are represented in intensity value. Photometric inversion from intensity to radiance is given in following sections in harmony with the notations of the existing works separately.

1. HDR Imaging Using Non-parametric Response Function

Debevec and Malik [34] estimate CRF assuming that the exposure durations of the images (Δt_j) , $j = 1, 2, \dots N$, are given. Denoting y_{ij} as the i th pixel of the j th image

$$y_{ij} = f(q_i \Delta t_j), \quad (1.11)$$

where f is the CRF, and q_i is the radiance value. Taking the inverse logarithm and substituting $g(\cdot) = \log f^{-1}(\cdot)$:

$$g(y_{ij}) = \log(q_i) + \log(\Delta t_j). \quad (1.12)$$

Both radiance values and CRF are unknown in Eq.1.12. Using the second derivative of the CRF as regularization term, cost function is given as

$$O = \sum_i \sum_j \{w(z_{ij})[g(y_{ij}) - \log(q_i) - \log(\Delta t_j)]\}^2 + \lambda \sum [w(z)g''(y)]^2, \quad (1.13)$$

Then radiance map is constructed by weighted average of radiance differences as such,

$$\log(q_i) = \frac{\sum_j w(y_{ij})[g(y_{ij}) - \log(\Delta t_j)]}{\sum_j w(y_{ij})} \quad (1.14)$$

2. HDR Imaging Using Polynomial Response Function

Mitsunga and Nayar [4] calculated CRF and radiance map by estimating exposure ratios instead of using exact exposure durations. Choosing one of the images

as reference and normalizing the exposure durations by the reference value, exposure ratio of the j th image becomes e_j . Similar to Eq.1.11, intensity value of the i th pixel of the j th image becomes

$$y_{ij} = f(q_i e_j). \quad (1.15)$$

substituting $g(\cdot) = f^{-1}(\cdot)$, irradiance values can be extracted by a polynomial as such

$$q_i e = g(y_i) = \sum_{k=0}^K c_k y^k, \quad (1.16)$$

where K denotes the order of polynomial. Exposure ratio can be extracted by dividing the irradiance values.

$$\frac{q_i e_j}{q_i e_{j+1}} = R_{j,j+1}, \quad (1.17)$$

where $R_{j,j+1}$ is the exposure ratio between image pair $(j, j+1)$. Cost function is expressed to minimize irradiance differences as such:

$$O = \sum_{j=1}^N \sum_{i=1}^P \left[\sum_{k=0}^K c_k y_{i,j}^k - R_{i,j+1} \sum_{k=0}^K c_k y_{i,j+1}^k \right]^2, \quad (1.18)$$

where N denotes the number of images and P the number of pixels in an image. This method is robust in the sense that it doesn't require exposure durations and estimate exposure ratios automatically starting from a rough estimate. However, this method raises u -ambiguity problem: multiple polynomials satisfying Eq.1.18. Also this method has a small tolerance to noise. Then HDR radiance map can be reconstructed by weighted averaging

$$q_i = \frac{\sum_{j=1}^N w(y_{i,j}) q_{i,j}}{\sum_{j=1}^N w(y_{i,j})}, \quad (1.19)$$

where $i = 1, 2, \dots, P$. The weighting function $w(\cdot)$ is chosen such as to give more credit to the region where sensitivity to radiance change is high. Sensitivity to radiance is achieved by dividing inverse camera response function by its derivative as follows

$$w(y_{i,j}) = \frac{g(y_{i,j})}{g'(y_{i,j})}. \quad (1.20)$$

3. HDR Imaging Using Parametric Response Function

Mann proposes multiple parametric CRF models in [38]. Mann constructs the photoquantity (q) value of a scene instead of radiance value. He defines q as the *quantity of light integrated over the spectral response of the particular camera system*. Most famous parametric CRF model is *preferred camera model*, which converts q to i th image (z_i) as

$$y_i = f(q) = \left(\frac{q^a}{q^a + 1} \right)^c, \quad (1.21)$$

where a and c are the camera parameters. Denoting k as the exposure ratio between images, it is easy to see that k directly acts on q . So, change in k results in an image with different intensity values (y_j) which can be related to the previous intensity value (y_i) through a tonal conversion function $g(\cdot)$ as

$$y_j = f(kq) = g[f(q)] = \frac{f(q).k^{ac}}{[\sqrt[c]{f(q).(k^a - 1) + 1}]^c} \quad (1.22)$$

Mann solved for camera parameters (a, c) using comparagram, which is the joint histogram of images (y_i, y_j). Camera parameters can be extracted from compragram by non-linear regression, such as Levenberg-Marquardt optimization.

Mann proposed using *certainty* function $\hat{c}(\cdot)$ to combine quantity values to from the ground truth quantity of the scene, which is similar to radiance map

$$\hat{q}(x, y) = \frac{\sum_i \frac{\hat{c}_i(q(x, y))}{k_i} f^{-1}(y_i(x, y))}{\sum_i \hat{c}_i(q(x, y))} \quad (1.23)$$

Certainty function is designed to give highest weight to the intensity value, which quickly responds to a quantity value.

$$\hat{c}_i(\log(q(x, y))) = \frac{d y_i(x, y)}{d \log \hat{q}(x, y)} \quad (1.24)$$

This method allows noise tolerance by thresholding the bins of comparagram. Although this method is computationally efficient, it is incapable of modeling arbitrary shaped CRFs.

D. Image Registration

Image registration methods can be grouped into three categories according to the motion model. An image can go through a global geometric transformation, such that the warping process of all pixels can be characterized by single matrix operation, i.e affine, perspective. We adopt global geometric transformation in this thesis since it is currently the most promising approach for real time SR implementation. Another registration method is using local geometric transformations, which is known as elastic or non-rigid transformation. In this technique, transformation parameters vary depending on the spatial location. Local transformation can be extracted by computing the geometric transformation of interest points in input and reference image. Last category is the optical flow, in which a dense motion field is computed for the whole image. In this approach, each pixel has its own motion vector. Optical flow techniques adopt regularization for motion vector estimation due to the ill posed nature that raises from aperture problem. Detailed survey of optical flow techniques is provided in [39].

Image registration is the process of aligning all images onto a common grid for global parametric motion model. Parametric motion can be estimated either minimizing the intensity error over the whole image or using the salient points. Thevenaz et al. proposed estimating the affine transformation between images minimizing the error using Levenberg-Marquardt optimization in a pyramidal approach [40]. Another approach for image registration is to use descriptive parts of images (features), instead of using all pixels. We adopt the feature based registration of Capel [41] for image alignment due to its robustness to illumination change for the scope of this thesis. Feature based image registration is a common method used for global image registration using the salient points. Automatic feature based image registration requires determining correspondences among images. Geometric transformation can be computed out of putative correspondences using robust estimators, i.e RANSAC. In

this thesis we adopt feature based registration by *planar projective transformation*, also known as *homography*.

Planar homography, also known as plane projective transformation, is a popular parametric motion model among the suggested numerous models. This model is capable of representing the perspective view, which is crucial as movement of pixel is actually a trajectory of 3-D motion. Projective transformation is the most generic form of 2-D coordinate transformations: translation, rigid, similarity and affine. Homographic transformation enables converting 3-D (homogenous) coordinates to 2-D (inhomogeneous) points [42]. This conversion is achieved by representing a homogeneous coordinate (x_1, x_2, x_3) as inhomogeneous point $(\frac{x_1}{x_3}, \frac{x_2}{x_3})$. Projective mapping can be achieved by multiplying homogenous coordinates with a 3×3 matrix as following:

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \quad (1.25)$$

Homographic coordinate conversion of a point is performed by denoting the k th pixel, which is (x_k, y_k) , as $(x_k, y_k, 1)$. Since we have 8 dof, h_{33} can be selected as 1. Applying the homography in Equation 1.25, transformed coordinates become

$$x_k' = \frac{h_{11}x_k + h_{12}y_k + h_{13}}{h_{31}x_k + h_{32}y_k + 1} \quad (1.26)$$

$$y_k' = \frac{h_{21}x_k + h_{22}y_k + h_{23}}{h_{31}x_k + h_{32}y_k + 1} \quad (1.27)$$

Rewriting Eq.1.26 and Eq. 1.27 for $k = 0, 1, 2, 3$ we get

$$\begin{bmatrix} x_0 & y_0 & 1 & 0 & 0 & 0 & -x_0x_0' & -y_0x_0' \\ x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1x_1' & -y_1x_1' \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2x_2' & -y_2x_2' \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3x_3' & -y_3x_3' \\ 0 & 0 & 0 & x_0 & y_0 & 1 & -x_0y_0' & -y_0y_0' \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1y_1' & -y_1y_1' \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2y_2' & -y_2y_2' \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -x_3y_3' & -y_3y_3' \end{bmatrix} \begin{pmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{pmatrix} = \begin{pmatrix} x_0' \\ x_1' \\ x_2' \\ x_3' \\ y_0' \\ y_1' \\ y_2' \\ y_3' \end{pmatrix} \quad (1.28)$$

Perspective motion parameters $h_{11} \dots h_{32}$ can be computed using pseudo-inverse.

Coordinate couples given in Equation 1.28 are putative matches among images to

be registered. These putative matches can be estimated from features of input images. Different type of features can be used to represent image characteristics. Points, edges, and contours are most commonly used as image features. Features should be invariant to condition changes in ambient. Harris corner detector is a favorite choice for feature detection [43]. Unfortunately Harris features are not capable of handling large illumination changes, and features may not repeat themselves under these circumstances. We propose a unique contrast invariant feature transform that produces interest points with high repeatability rate in Chapter II.

Features are the distinctive points in an image where the variance highly varies with respect to its neighborhood. These salient points are used in homography estimation since they have a tendency to occur with high frequency under different geometric and photometric conditions. Detailed literature survey for feature detection is provided in Chapter II.

Two images are given in Figure 7 to illustrate the registration of photometrically and geometrically diverse images. Input and reference images are given in Figure 7 (a) and (b), and invariant features for these images are shown in Figure 7 (c) and (d), respectively.

There are four main steps in homography estimation:

1. Feature extraction
2. Feature matching by outlier rejection
3. Sub-pixel refinement of matched features
4. Random Sample Consensus (RANSAC)

Features in corresponding images can be matched using any similarity measure. A window is selected around a feature to measure its similarity to other features. Feature matching can be done either minimizing the error (i.e MSE) or maximizing the correlation of feature blocks. Normalized cross correlation (NCC) is the most

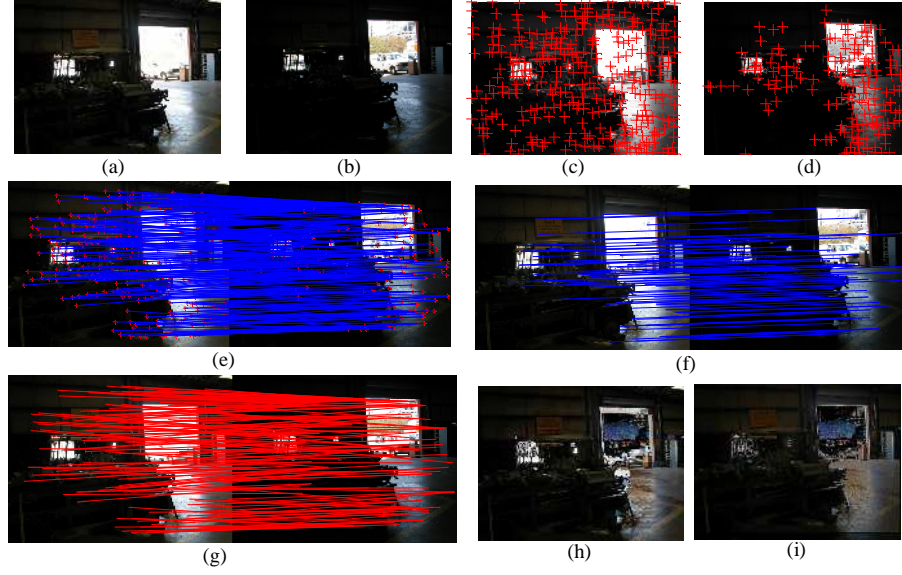


Fig. 7.: Feature based homography estimation. (a) Input image (b) Reference Image (c) Contrast invariant features of input image (d) Contrast invariant features of reference image (e) Putative matches using normalized cross correlation (NCC) (f) Inliers estimated using RANSAC (g) Outliers estimated using RANSAC (h) Residual of input and reference image without registration (i) Residual of input and reference image after registration

suitable similarity metric to compensate for the photometric change. NCC achieves this by subtracting the mean values before correlating blocks. This compensates for the exposure changes as illumination changes directly act on the mean value of the blocks. To compute the NCC of features in input I_1 and reference I_2 image we take a local window with same size as Ω_1 and Ω_2 , respectively. Denoting the local mean of images as \bar{I}_1 and \bar{I}_2 within the feature neighborhood Ω_1 and Ω_2 , NCC of features can be computed as following:

$$\text{NCC}(I_1, I_2, \Omega_1, \Omega_2) = \frac{\sum_{\mathbf{x}_1 \in \Omega_1} \sum_{\mathbf{x}_2 \in \Omega_2} (I_1(\mathbf{x}_1) - \bar{I}_1) \odot (I_2(\mathbf{x}_2) - \bar{I}_2)}{(\sum_{\mathbf{x}_1 \in \Omega_1} \sum_{\mathbf{x}_2 \in \Omega_2} (I_1(\mathbf{x}_1) - \bar{I}_1)) (\sum_{\mathbf{x}_1 \in \Omega_1} \sum_{\mathbf{x}_2 \in \Omega_2} (I_2(\mathbf{x}_2) - \bar{I}_2))}, \quad (1.29)$$

where \mathbf{x}_1 and \mathbf{x}_2 are spatial location vectors used to span the local window Ω_1 and Ω_2 , and \odot denotes element by element multiplication.

Correlated features contains both true matches (inliers) and false matches (outliers). A match is considered to be valid if similarity score (NCC) of corresponding features passes a threshold. Features having NCC scores below a threshold are consid-

ered to be outliers. Eliminating the feature matches with low NCC scores we obtain putative matches. These putative matches might still contain outliers that contradict with the actual geometric transformation. However, elimination depending on NCC scores yield a reasonable starting point. RANSAC [44] can be used to eliminate outliers and estimate homography out of putative matches.

Feature points are aimed to be classified as inliers and outliers in RANSAC. Algorithm is repeated for certain number of times and iteration returning the highest number of elements in inlier class is chosen as the winner. In each iteration 4 number of points are selected for homography computation. Homography is estimated using the coordinates of these randomly selected features using Equation 1.28. All feature points in input image are warped onto the reference image using this estimated homography. A feature is considered to be inlier if its warped location is within a distance to the spatial location of its putative match.

Constraining the feature matching in a window increases the efficiency of the method since searching a match all over the image would be exhaustive. Corner matching procedure is depicted in Figure 7 of Chapter I . Putative matches are given in Figure 7 (e), which is composed of inliers and outliers. Inliers and outliers are shown explicitly in Figure 7 (f) and (g), correspondingly. Difference images before and after registration are shown in Figure 7 (h) and (i), respectively.

CHAPTER II

ILLUMINATION ROBUST INTEREST POINT DETECTION

Most interest point detection algorithms are highly sensitive to illumination variations. In this chapter, we present a method to find interest points robustly even under large photometric changes. The method, which we call contrast invariant feature transform (CIFT), determines salient interest points in an image by calculating and analyzing *contrast signatures*. A contrast signature shows the response of an interest point detector with respect to a set of contrast stretching functions. The CIFT is generic and can be used with most interest point detectors. We demonstrate that the CIFT improves the repeatability rate of the Harris corner detector significantly (by around 25% on average in the experiments).

A. Introduction

Interest point detection is a necessary task in a variety of computer vision applications, including object recognition, tracking, stereo imaging, and mosaic construction. Although the term interest point detector is interchangeably used with the term corner detector; interest points do not necessarily correspond to physical corner points. The performance of an interest point detector is evaluated in terms of its accuracy and consistency. An interest point detector should have good spatial localization and be robust against noise and geometric/photometric changes.

One of the earliest interest point detection algorithms is the Moravec corner detector [45]. In the algorithm, a patch around a pixel is taken and compared with the neighboring patches. If the pixel is on a smooth region or an edge, there would be at least one neighboring patch that is very similar. For a corner point, all neighboring patches would be different. Therefore, the corner strength at a pixel is defined as the smallest sum of squared differences between the center patch and its neighboring patches. The problem with this algorithm is that only a finite number of neighboring

patches, (patches in horizontal, vertical, and diagonal directions), are considered; hence, the algorithm is not isotropic.

Harris and Stephens [43] derived a better corner detector by applying Taylor series expansion to the sum of squared difference between a patch and its arbitrarily shifted version. This expansion yields a 2×2 autocorrelation matrix, whose eigenvalues indicates the magnitudes of two orthogonal gradient vectors. For a corner point, both magnitudes should be large. Since the calculation of eigenvalues and eigenvectors is costly, a corner strength measure based on the determinant and trace of the autocorrelation matrix is proposed. The local maxima of the corner strength map are chosen as the corner points. This method is known as the Harris or the Plessey corner detector in the literature and it is probably the most popular interest point detector.

Another important interest point detector is the SUSAN corner detector [46]. The SUSAN operator finds corners based on the fraction of pixels that are similar to the center pixel within a small circular region. A low value of this fraction indicates a corner. The corneriness response at a pixel is basically obtained by subtracting the number of similar pixels from a fixed geometric threshold. A non-maxima suppression operation determines the corner points. The SUSAN operator is highly robust to noise. Modifications of the SUSAN operator are used for edge detection and for image denoising. (The well-known Bilateral filter is indeed same as the SUSAN image denoising filter.)

In [47], phase congruency is used to detect edges and corners. The method is based on the *local energy model*, which postulates that the frequency components are in phase at corners and edges [48]. One advantage of this method is the relative insensitivity against illumination changes. The phase congruency method also has very accurate spatial localization.

Recently, methods to extract scale and affine invariant interest points have been

proposed. A review of these methods can be found in [49]. A comprehensive evaluation of the interest point detectors is provided in [50]. Among these methods, the scale invariant feature transform (SIFT) [51] method also presents a region descriptor based on the local histogram of the gradient vectors to achieve illumination invariance in addition to scale invariance. In the illumination invariance experiments of [50], the SIFT method is among the best performers.

In this paper, we present a method, contrast invariant feature transform (CIFT), to improve the robustness of feature detection against illumination changes. The CIFT stretches the histogram of an image around a set of pixel intensities; an interest point detector is applied to each contrast-stretched image to produce a three-dimensional cornerness map. Interest points are determined from this map.

The CIFT is generic and can be combined with a feature detector that creates feature strength map and then applies non-maxima suppression to it. In this paper, we show how the CIFT improves the repeatability rate of the Harris corner detector. Section B reminds the standard Harris corner detector. Section C explains the idea behind the CIFT and also presents its application on the Harris corner detector. Section E compares several interest points detectors experimentally. Section F concludes the paper.

B. Harris Corner Detector

One of the most commonly used interest point detectors in computer vision applications is the Harris corner detector [43]. The Harris corner detector is based on the autocorrelation matrix of the image gradients. The autocorrelation matrix $\mathbf{A}(x, y)$ at a pixel (x, y) is given as follows:

$$\mathbf{A}(x, y) = \begin{pmatrix} \sum_{(m,n) \in N} \left(\frac{\partial}{\partial x} I(m, n) \right)^2 & \sum_{(m,n) \in N} \frac{\partial}{\partial x} I(m, n) \frac{\partial}{\partial y} I(m, n) \\ \sum_{(m,n) \in N} \frac{\partial}{\partial x} I(m, n) \frac{\partial}{\partial y} I(m, n) & \sum_{(m,n) \in N} \left(\frac{\partial}{\partial y} I(m, n) \right)^2 \end{pmatrix}, \quad (2.1)$$

where $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ calculates the gradients in horizontal and vertical directions, respectively; and N is a set of pixels around (x, y) . Usually a weighted sum of the gradients within N is taken using a Gaussian function centered at (x, y) to give more weight to the pixels that are close to (x, y) . Instead of computing the eigenvalues of $\mathbf{A}(x, y)$, [43] proposes a cornerness response function $R(x, y)$ that can be computed efficiently with the determinant and the trace of the matrix $\mathbf{A}(x, y)$:

$$R(x, y) = \det \mathbf{A}(x, y) - k(\text{trace} \mathbf{A}(x, y))^2, \quad (2.2)$$

where k is a small positive constant controlling the cornerness sensitivity of the detector. After calculating the cornerness response for all pixels, non-maxima suppression is used to get the corner points.

C. Contrast Invariant Feature Transform

The CIFT is a method to improve illumination invariance of feature detectors. The underlying idea is to stretch the image contrast as a function of intensity to span the space of possible photometric transformations. By applying a feature detector to a contrast-stretched image, the response of the feature detector under a particular illumination condition is simulated. The collection of the responses under a set of illumination conditions forms a signature for each pixel. The signatures can then be used to characterize a pixel and to find illumination-invariant interest points.

The contrast stretching function that we use in our experiments is the sigmoid function, which has the following form:

$$f_c(I) = \frac{1}{1 + e^{-\gamma(I-c)}}, \quad (2.3)$$

where I is the normalized intensity value in the range of $[0, 1]$, c is the contrast center around which the contrast is stretched, and γ determines the slope of the sigmoid function. Figure 8 illustrates the sigmoid function.

By applying the contrast stretching function at different contrast centers, we

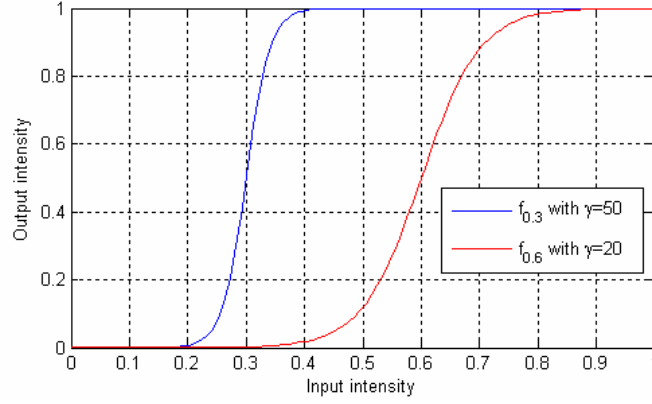


Fig. 8.: Sigmoid functions at contrast centers $c = 0.3$ and $c = 0.6$ are plotted. The parameter γ controls the slope of the sigmoid.

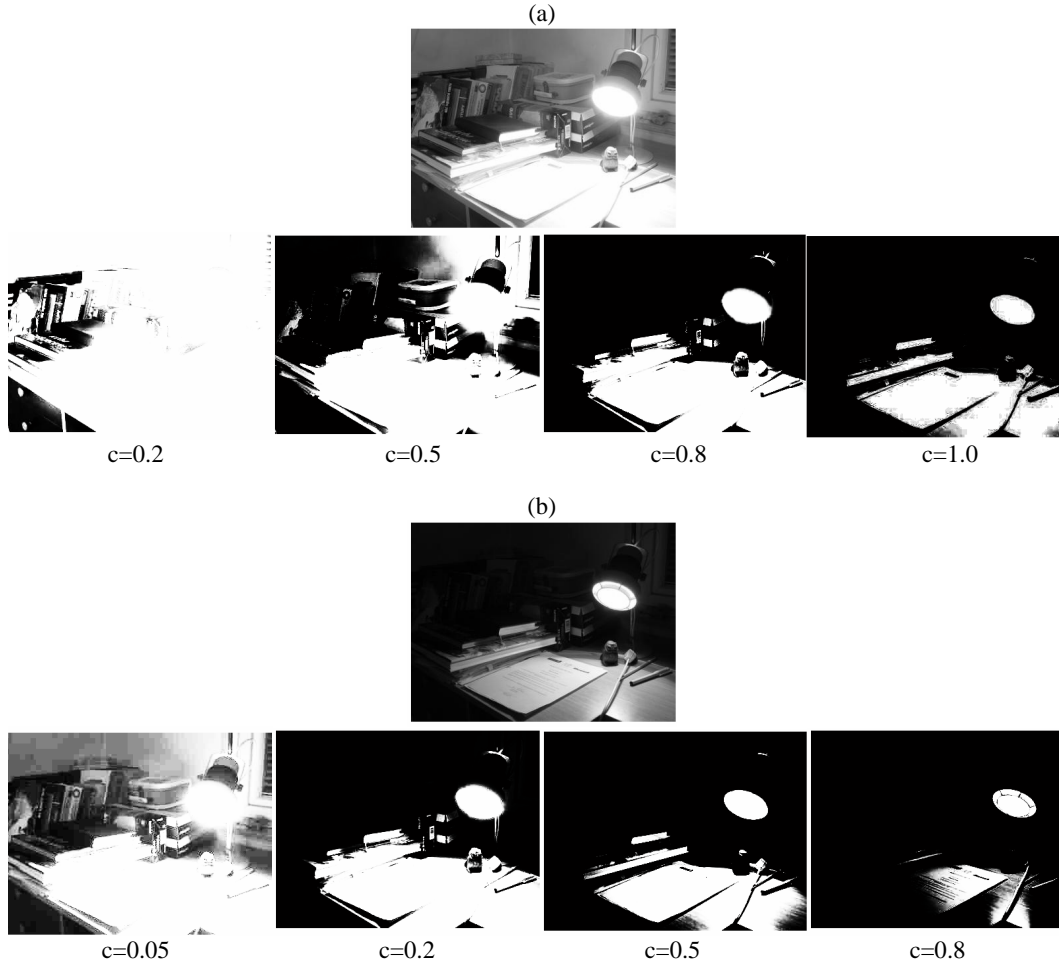


Fig. 9.: Contrast stretched versions of two images are displayed at several centers c . ($\gamma = 50$ for all.)

obtain a set of contrast-stretched images. (Figure 9 shows a set of contrast-stretched versions of two images at various contrast centers. Notice that two different images produces similar responses at particular contrast centers. For example, $c = 0.8$ for Image (a) produces an image that is very similar to the output of $c = 0.2$ for Image (b). Likewise, the output of $c = 1.0$ for Image (a) is similar the output of $c = 0.5$ for Image (a). Also, notice how some details in $c = 0.05$ for Image (b) get apparent as in Image (a).) A feature detector is then applied to each of these contrast-stretched images. In case of the Harris corner detector, first, the autocorrelation matrix at each pixel is found:

$$\mathbf{A}(x, y; c) = \begin{pmatrix} \sum_{(m,n) \in N} \left[\frac{\partial}{\partial x} I_c(m, n) \right]^2 & \sum_{(m,n) \in N} \left[\frac{\partial}{\partial x} I_c(m, n) \right] \left[\frac{\partial}{\partial y} I_c(m, n) \right] \\ \sum_{(m,n) \in N} \left[\frac{\partial}{\partial x} I_c(m, n) \right] \left[\frac{\partial}{\partial y} I_c(m, n) \right] & \sum_{(m,n) \in N} \left[\frac{\partial}{\partial y} I_c(m, n) \right]^2 \end{pmatrix}, \quad (2.4)$$

where $I_c(m, n) \equiv f_c(I(m, n))$ is the contrast-stretched image at contrast center c . And then, the cornerness response is calculated:

$$R(x, y; c) = \text{trace} \mathbf{A}(x, y; c) - k(\text{trace} \mathbf{A}(x, y; c))^2. \quad (2.5)$$

Note that the cornerness response is a function of the contrast center c . $R(x, y; c)$ for all $(x, y; c)$ is a three-dimensional matrix that can be used to analyze the pixels in terms of their cornerness. By plotting $R(x, y; c)$ as a function of c at a pixel position, we can obtain the “contrast signature” of that pixel. Figure 10 shows the contrast signatures at several pixels on an image. Five pixels are selected from the image. These pixels are enumerated from 1 to 5. Among these pixels, pixel 1 is on a corner with the highest contrast. When we look at the contrast signature of that pixel, we see that it has a strong response over a wide range of c . On the other hand, a corner pixel with low contrast returns a large response over a limited range of c . (Although not plotted in the figure, pixels that are on edges and smooth regions return a low response for all values of c .)

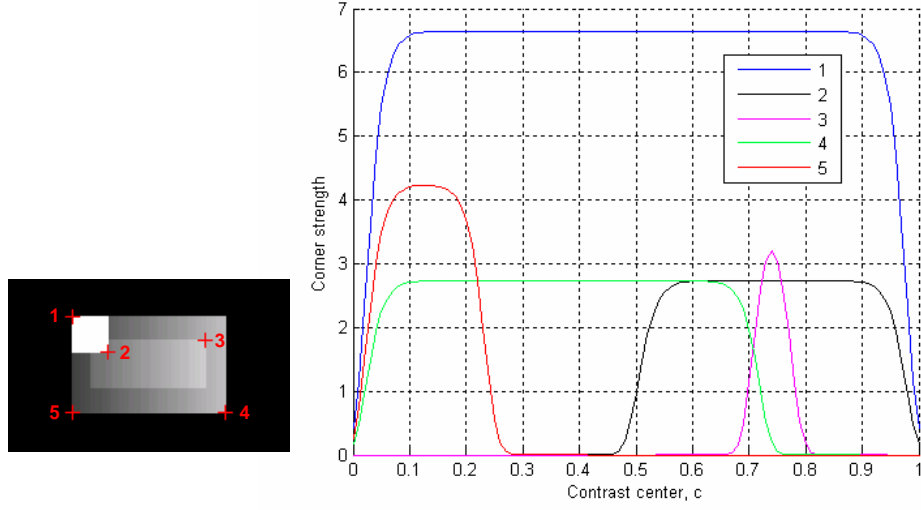


Fig. 10.: Sample pixels and the corresponding contrast signatures are shown.

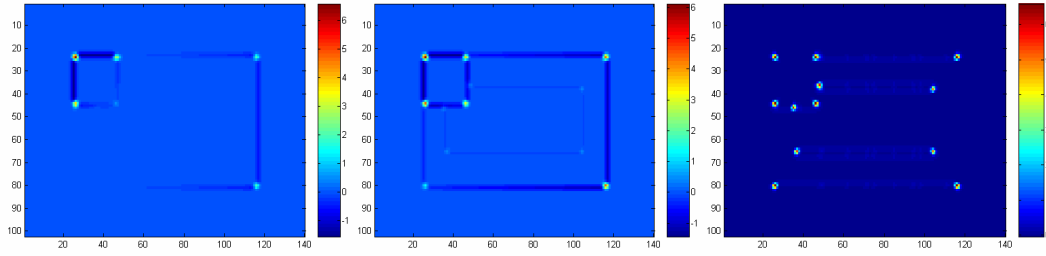


Fig. 11.: Left to right: Cornerness responses of the standard Harris corner detector $R(x, y)$, CIFT Harris with $R_{\text{area}}(x, y)$, and CIFT Harris with $R_{\text{max}}(x, y)$ are displayed for the image in Figure 2.

We can define different cornerness measures based on the contrast signatures. Two possible cornerness measures are the area and the peak value of the signatures:

$$R_{\text{area}}(x, y) = \int_0^1 R(x, y; c) dc \quad (2.6)$$

and

$$R_{\text{max}}(x, y) = \max_c R(x, y; c). \quad (2.7)$$

Figure 11 shows these cornerness responses for the image given in Figure 10. As

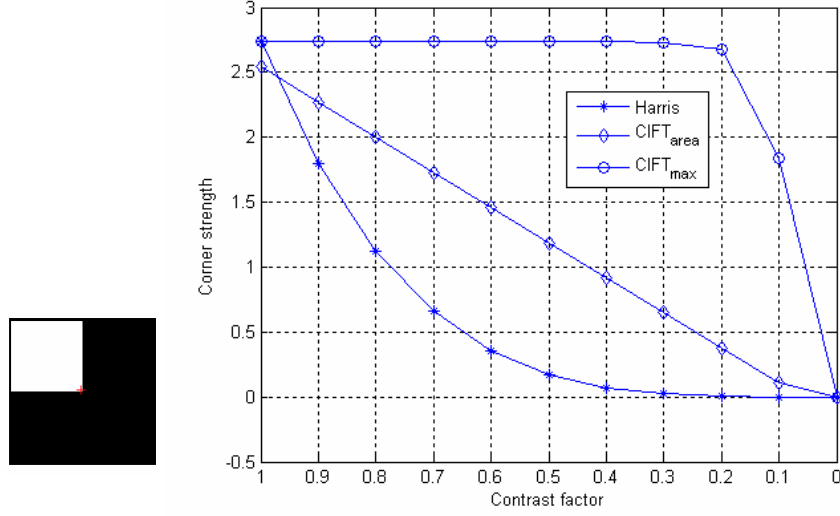


Fig. 12.: The intensity of the white patch is reduced from 1 to 0. The corneriness responses of the standard Harris corner detector $R(x, y)$, CIFT Harris with $R_{\text{area}}(x, y)$, and CIFT Harris with $R_{\text{max}}(x, y)$ as a function of this intensity are plotted for the marked pixel.

seen, the standard Harris corner detector misses the corners with low contrast; on the other hand, $R_{\text{area}}(x, y)$ and $R_{\text{max}}(x, y)$ determines all the corners well. Among all, $R_{\text{max}}(x, y)$ has the best distinguishing response.

To test how the corneriness response changes as function of contrast, we set up the experiment shown in Figure 12. The intensity of the white patch in a black image is reduced from 1 to 0; and the responses $R(x, y)$, $R_{\text{area}}(x, y)$, and $R_{\text{max}}(x, y)$ are measured at the corner of the white patch. It is observed that the standard Harris response drops sharply, while the responses based on the contrast signature are more robust.

D. Experiments

In this section, we report the experiments demonstrating how the CIFT improves the repeatability rate of the Harris corner detector under large photometric changes and also compare with several state-of-the-art corner detectors under the same conditions.

There are three data sets, which are shown in figures 13, 14, and 15. There are

non-uniform illumination changes and severe saturation in these images. In each set, one image is chosen as the reference image, and the repeatability rate is computed for each of the remaining images.

There are different measures of repeatability rate for interest point detectors; we use the following measure. Suppose N_1 is the number of interest points in image 1; N_2 is the number of interest points in image 2; and N is the number of common interest points. Then, the repeatability rate is defined as $N/\min(N_1, N_2)$. An interest point in an image is repeated if there is an interest point within a 3×3 neighborhood of the other image.

In the experiments, we used the following parameters. For the Harris corner detector, the value of k was 0.04. In calculating the autocorrelation matrix, we used a Gaussian function with standard deviation of 1 to get the weighted sum of the gradients within a 7×7 local neighborhood. For the non-maxima suppression, the pixels with cornerness response less than a threshold were eliminated, and the local maxima within 3×3 regions were found for the remaining pixels. The threshold value, below which the corresponding pixels are eliminated, was set to 2% of the maximum value of the cornerness response. For the CIFT Harris algorithm, the signatures were obtained for $c \in [0, 1]$ with step sizes of 0.05 and with $\gamma = 50$. These parameters were decided after some trial-and-error to have a good overall performance. In addition to the standard Harris and the CIFT Harris, we included results of the Phase Congruency [47], SIFT [51], and SUSAN [46] methods. The softwares for these methods are available online from the original authors. For these methods, the parameters were chosen as the default values given in these softwares.

As mentioned earlier, we have considered two possible CIFT approaches; in the CIFT Harris with R_{area} approach, the area under a signature is used as the cornerness measure; in the CIFT Harris with R_{max} approach, the maximum value of the signature is used as the cornerness measure. During our experiments, we have noticed that,

for real images, the R_{area} approach yields higher repeatability rate than the R_{max} approach on average. (One possible explanation is that noise which is amplified as a result of contrast stretching returns a high value for a limited range of c . However, by taking the maximum of the signature, it is mistaken as a corner point.) Since the R_{area} approach is better than the R_{max} for real images, we report the results for the R_{area} approach only.

The results are given in Figure 16. In the figure, both the repeatability rate and the number of repeated corners are plotted. As seen in the results, the CIFT improves the repeatability rate of the standard Harris corner detector by around 25% on average. The improvement is larger for images with large photometric differences. When the number of repeated corners is examined, it can be seen that the number of repeated corners is significantly larger for the CIFT Harris compared to the standard Harris. The number of repeated corners might be important in some applications, for example, when registering images with relatively small overlapping areas. Also, one might try to improve the repeatability rate by increasing the threshold of the non-maxima suppression or by ranking the pixels according to the corneriness strength and then taking the top, say, M of them. For these scenarios, the CIFT Harris has larger space for improvement than the standard Harris.

While improving the standard Harris, the CIFT Harris beats the other interest point detectors as well. In all experiments, the CIFT Harris has higher repeatability rate than the Phase Congruency, SIFT, and SUSAN methods.

E. Conclusions

In this chapter, we proposed a method, contrast invariant feature transform (CIFT), to detect interest points robustly under varying illumination conditions and demonstrated how it improves the repeatability rate of the standard Harris corner detector. The CIFT Harris also performs better than several other state-of-the-art corner detection algorithms. An important drawback of the CIFT method is the

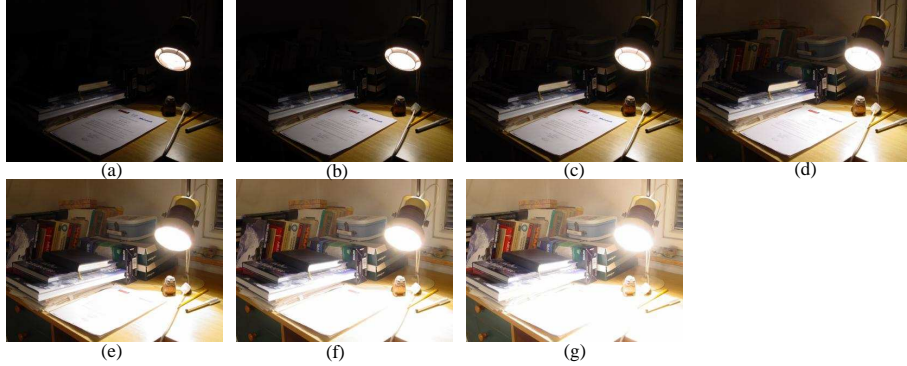


Fig. 13.: Data set 1. Image in (d) is chosen as reference. Repeatability rates between this image and the other images are computed. The images in (a),(b),(c),(e),(f), and (g) are labeled as 1 to 6 in Figure 16. The image size is 300×400 .

computational complexity. In the current implementation, if the contrast center c is sampled at m points, then the computational complexity of the CIFT is approximately m times the computational complexity of the method it is applied. As a future work, we are looking into possible ways to reduce the complexity. One possible approach is to reduce γ and sample c less densely. Of course, for applications where performance is more important than computational complexity, the CIFT would be very beneficial. A second possible future work is to apply the CIFT on affine invariant interest point detectors to obtain affine and photometric invariant feature detectors. In the paper, we have presented two cornerness measures based on the contrast signature. A third possible future work is to define and test alternative cornerness measures to achieve higher repeatability rates. For example, one might look at the contrast signature area within a certain range of the peak value as the cornerness measure. Lastly, the method can be combined with the scale invariant region descriptors to obtain an interest point detector that is reliable under photometric and geometric variations. One approach could be to photometrically normalize the local region around a point based on the contrast signatures, *i.e.*, $\int \bar{R}(x, y; c) f_c(I(x, y)) dc$, where $\bar{R}(x, y; c) = R(x, y; c) / \int_c R(x, y; c) dc$ is the area normalized signature; and then apply, for example, the SIFT descriptor to these *illumination normalized* images. (See Figure 17 for an example of such illumination normalization.).

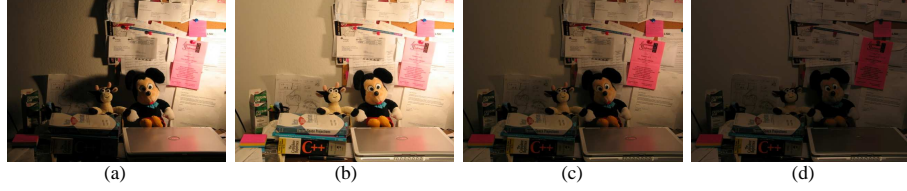


Fig. 14.: Data set 2. Image in (a) is chosen as the reference image. Repeatability rates between this image and the other images are computed. The images in (b),(c), and (d) are labeled as 1,2, and 3 in Figure 16. The image size is 480×640 .

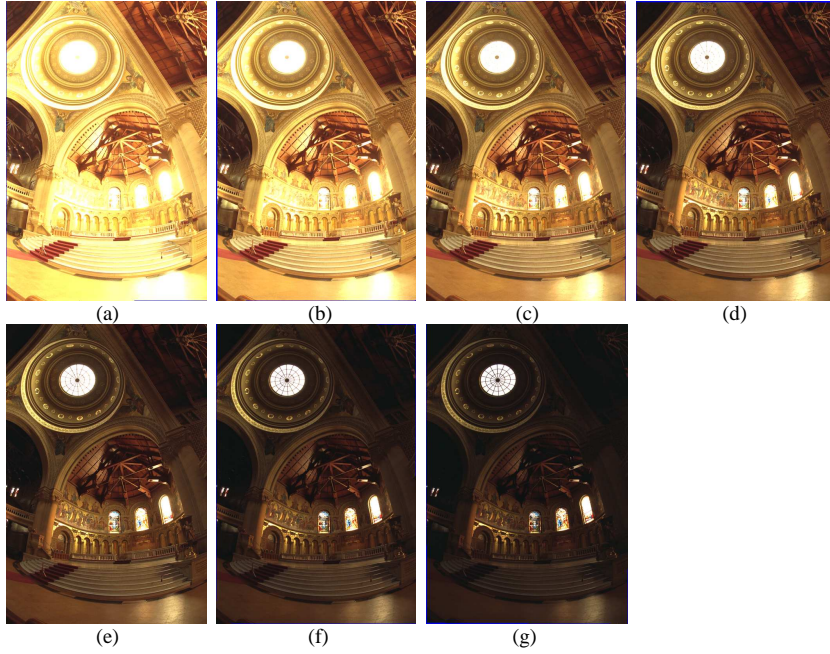


Fig. 15.: Data set 3. Image in (d) is chosen as reference. Repeatability rates between this image and the other images are computed. The images in (a),(b),(c),(e),(f), and (g) are labeled as 1 to 6 in Figure 16. The image size is 768×512 .

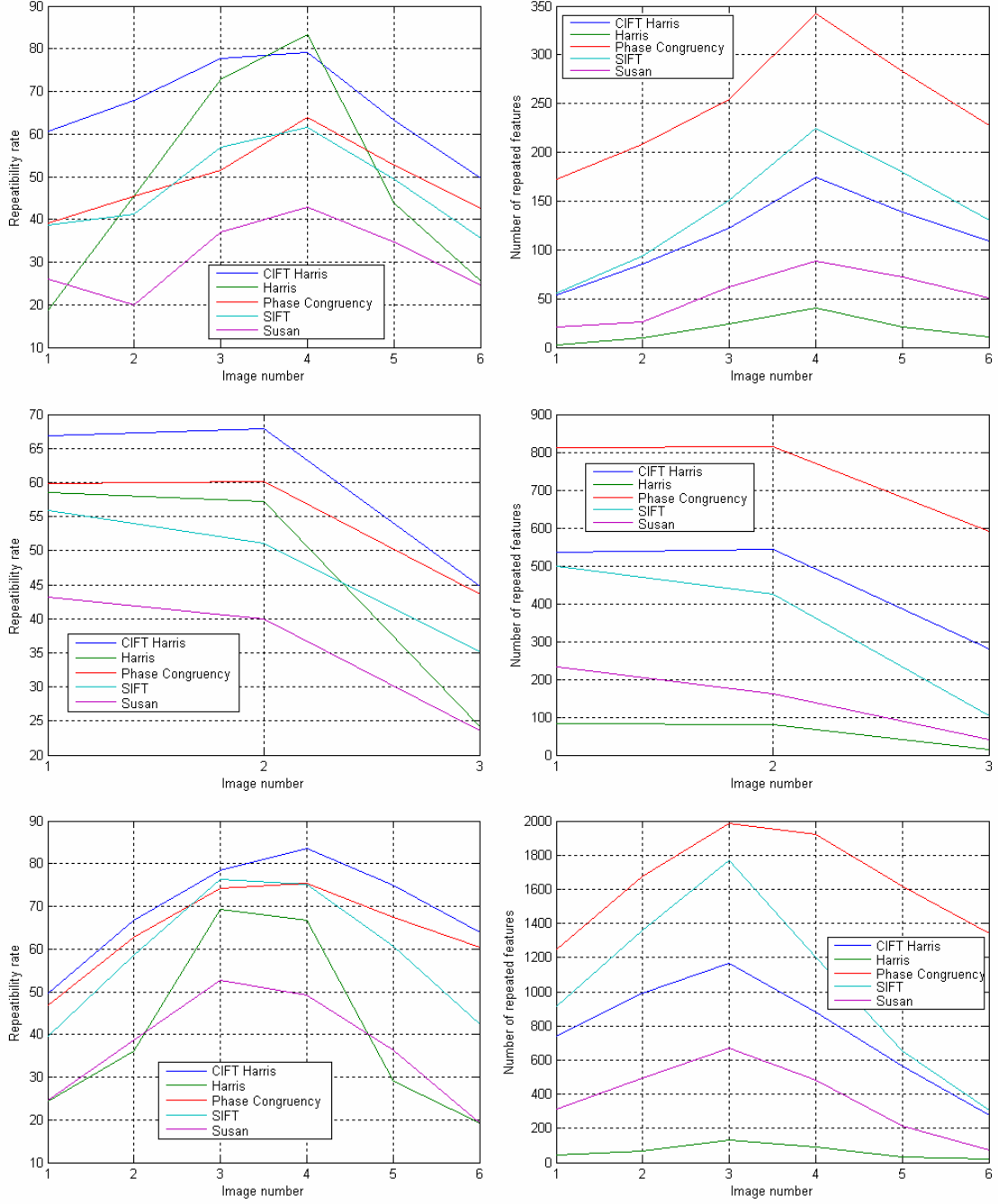


Fig. 16.: Top to bottom: Experimental results (the repeatability rates and the number of repeated corners) for data set 1, data set 2, and data set 3, respectively.



Fig. 17.: Upper row: Images at two different illumination conditions. The center pixels are marked. Lower row: Corresponding *illumination normalized* images based on the contrast signatures of the center pixels.

CHAPTER III

SUPER RESOLUTION UNDER PHOTOMETRIC DIVERSITY OF IMAGES

Super resolution (SR) is a well-known technique to increase the quality of an image using multiple overlapping pictures of a scene. SR requires accurate registration of the images, both geometrically and photometrically. Most of the SR articles in the literature have considered geometric registration only, assuming that images are captured under the same photometric conditions. This is not necessarily true as external illumination conditions and/or camera parameters (such as exposure time, aperture size and white balancing) may vary for different input images. Therefore, photometric modeling is a necessary task for super resolution. In this chapter, we investigate super-resolution image reconstruction when there is photometric variation among input images.

A. Introduction

In this chapter, we focus on a new issue in SR: How to perform SR when some of the input images are photometrically different than the others. Other than a few recent papers, almost all SR algorithms in the literature assume that input images are captured under the same photometric conditions. This is not necessarily true in general. External illumination conditions may not be identical for each image. Images may be captured using different cameras that have different radiometric response curves and settings (such as exposure time and ISO settings). Even if the same camera is used for all images, camera parameters (exposure time, aperture size, white balancing, gain, etc.) may differ from one image to another. (Almost all modern cameras have automatic control units adjusting the camera parameters. Low-end “point-and-shoot” digital cameras determine these parameters based on some built-in algorithms and do not allow users to change them. A slight repositioning of the camera or a change in the scene may result in a different set of parameters.) Therefore,

an SR algorithm should include a photometric model as well as a geometric model and incorporate these models in the reconstruction.

For accurate photometric modeling, the camera response function (CRF) and the photometric camera settings should be taken into account. The CRF, which is the mapping between the irradiance at a pixel to the output intensity, is not necessarily linear. Charges created at a pixel site due to incoming photons may exceed the holding capacity of that site. When the amount of charge at a pixel site approaches the saturation level, the response may deviate from a linear response. When a pixel site saturates, it outputs the same intensity even if more photons come in. (If photons keep coming after saturation, the charge starts to fill the neighboring pixels unless there is an anti-blooming technology in the sensor.) In addition, camera manufacturers may also introduce intentional nonlinearity to CRF to improve contrast and visual quality.

The CRF saturation and the finite number of bits (typically eight bits per channel) to represent a pixel intensity limit the resolution and the extend of the dynamic range that can be captured by a digital camera. Because a real scene typically has much wider dynamic range than a camera can capture, an image captures only a limited portion of the scene’s dynamic range. By changing exposure rate, it is possible to get information from different parts of a scene. In high-dynamic-range (HDR) imaging research, multiple low-dynamic-range (LDR) images (that are captured with different exposure rates) are combined to produce a HDR image [52, 34, 53, 54]. This process requires estimation or knowledge of the exposure rates and CRF. Spatial registration, lens flare and ghost removal, vignetting correction, compression and display of HDR images are some of the other challenges in HDR imaging.

Despite the likelihood of photometric variations among images of a scene, there are few SR papers addressing reconstruction with such image sets. In [55, 41], photometric changes were modeled as global gain and offset parameters among image intensities. This is a successful model when photometric changes are small. When

photometric changes are large, nonlinearity of CRF should be taken into consideration. In [56], we included a nonlinear CRF model in the imaging process, and proposed an SR algorithm based on the maximum a posteriori probability estimation technique. The algorithm produces the high-resolution irradiance of the scene; it requires estimation of the CRF and its inverse explicitly. The algorithm derives a specific *certainty function* using the Taylor series expansion of the inverse of the CRF. (As we will see, certainty function controls the contribution of each pixel in reconstruction. It gives less weight to noisy and saturated pixels than reliable pixels. It is necessary for a good reconstruction performance.)

We propose an alternative method. The method works in the intensity domain instead of the irradiance domain as proposed in [56]. It is not necessary to estimate the CRF or the camera settings; intensity-to-intensity mapping is sufficient. The spatial resolution of the reference image is enhanced without going to the irradiance domain. In addition, the photometric weight function is generic in the derivations; no Taylor series expansion is required.

The rest of the chapter is as follows. In Section B, we compare two photometric models that have been applied in SR. We show that nonlinear photometric modeling is necessary when photometric changes are significant. We then investigate two possible approaches for SR under photometric diversity in Section C. In Section D, we explain our approach to accurate geometric and photometric registration. We provide experimental results with real data sets in Section E. Conclusions and future work are given in Section F.

B. Photometric Modeling

For a complete SR algorithm, spatial and photometric processes of an imaging system should be modeled. Spatial processes (spatial motion, sampling, point spread function) have been investigated relatively well; here, we investigate photometric modeling. As mentioned earlier, in the context of SR, two photometric models have

been used. The first one is the affine model used in [55, 41], and the second one is the nonlinear model used in [56]. In this section, we review and compare these two models.

1. Affine Photometric Model

Suppose that N images of a static scene are captured and these images are geometrically registered. Let \mathbf{q} be the irradiance of the scene, and \mathbf{z}_i be the i th measured image.¹ According to the affine model, the relation between the irradiance and the image is as follows:

$$\mathbf{z}_i = a_i \mathbf{q} + b_i, \quad i = 1, \dots, N, \quad (3.1)$$

where the gain (a_i) and offset (b_i) parameters can model a variety of things, including global external illumination changes and camera parameters such as gain, exposure rate, aperture size, and white balancing. (In HDR image construction from multiple exposures, only the exposure rate is manually changed, keeping the rest of the camera parameters fixed [52]. In such a case, the offset term can be neglected.) Then, the i th and the j th images are related to each other as follows:

$$\mathbf{z}_j = a_j \mathbf{q} + b_j = a_j \left(\frac{\mathbf{z}_i - b_i}{a_i} \right) + b_j = \frac{a_j}{a_i} \mathbf{z}_i + \frac{a_i b_j - a_j b_i}{a_i}. \quad (3.2)$$

Defining $\alpha_{ji} \equiv \frac{a_j}{a_i}$ and $\beta_{ji} \equiv \frac{a_i b_j - a_j b_i}{a_i}$, we can in short write (3.2) as

$$\mathbf{z}_j = \alpha_{ji} \mathbf{z}_i + \beta_{ji}. \quad (3.3)$$

The affine relation given in (3.3) is used in [41] to model photometric changes among the images to be used in SR reconstruction. In [41], the images are first geometrically registered to the reference image to be enhanced. (A feature-based registration method is used. Corner points in the images are extracted and matched using normalized cross correlation. Perspective registration parameters are estimated

¹In our formulations, images are represented as column vectors.

after outlier rejection.) After geometric registration the relative gain and offset terms with respect to the reference image are calculated with least squares estimation. Each image is photometrically corrected using the gain and offset terms. This is followed by SR reconstruction.

Although the affine transformation in (3) can handle small photometric changes, the conversion accuracy decreases drastically in case of large changes. This is why in HDR imaging nonlinear photometric modeling is preferred over affine modeling.

2. Nonlinear Photometric Model

A typical image sensor has a nonlinear response to amount of light it receives. Estimation of nonlinear camera response function (CRF) becomes crucial in a variety of applications, including HDR imaging, panoramic image construction [57, 58], photometric stereo [59], bidirectional reflectance distribution function (BRDF) estimation, and thermography.

According to the nonlinear photometric model, an image \mathbf{z}_i is related to the irradiance \mathbf{q} of the scene as follows:

$$\mathbf{z}_i = f(a_i \mathbf{q} + b_i), \quad (3.4)$$

where $f(\cdot)$ is the camera response function (CRF), and a_i and b_i are again the gain and offset parameters as in (3.1). Then, two images are related to each other as follows:

$$\mathbf{z}_j = f\left(\frac{a_j}{a_i} f^{-1}(\mathbf{z}_i) + \frac{a_i b_j - a_j b_i}{a_i}\right) = f(\alpha_{ji} f^{-1}(\mathbf{z}_i) + \beta_{ji}). \quad (3.5)$$

The function $f(\alpha_{ji} f^{-1}(\cdot) + \beta_{ji})$ is known as the intensity mapping function (IMF). (Note that in some papers such as [60], the offset term in the above equation is neglected and the term $f(\alpha_{ji} f^{-1}(\cdot))$ is called the IMF.) Although IMF can be constructed using CRF and exposure ratios, it is not necessary to estimate camera parameters to find IMF. IMF can be extracted directly from the histograms of the images [60]. Another way to estimate IMF is proposed in [61], which estimates

IMF from two-dimensional intensity distribution of input images. Slenderizing this joint intensity distribution results in IMF. [61] also estimates the CRF and exposure rates using a nonlinear optimization technique. CRF can also be estimated without finding IMF. In [62], a parametric CRF model is proposed; and these parameters are estimated iteratively. [63] uses a polynomial model instead of a parametric model. In [34], a nonparametric CRF estimation technique with a regularization term is presented. Another nonparametric CRF estimation method is proposed in [35], which also includes modeling of noise characteristics.

3. Comparison of Photometric Models

Here, we provide an example to compare affine and nonlinear photometric models. In Figure 1(a, b, c, d), we provide four images captured with a hand-held digital camera. One of the images is set as the reference image (Figure 1(d)) and the others are converted to it photometrically using the affine and nonlinear models. (Before photometric conversion, images were registered geometrically.) The residual images computed using the affine model (Figure 1(e, f, g)) and the nonlinear model (Figure 1(i, j, k)) are displayed. The affine model parameters are estimated using the least squares technique and are shown in Figure 1(h). The nonlinear IMFs are estimated using the method in [38]. The estimated mappings are shown in Figure 1(l). As seen from the residual images, the nonlinear model works better than the affine model. The affine model performs well when the exposure ratios are close; the model becomes more and more insufficient as the exposure ratios differ more. Figure 19 demonstrates this for a larger set of exposure ratios, ranging from 2 to 50.

A super-resolution algorithm requires an accurate modeling of the imaging process. The restored image should be consistent with the observations given the imaging model. A typical iterative SR algorithm (POCS-based [27], Bayesian [18], iterated back projection [11]) starts with an initial estimate, calculates an observation using the imaging model, finds the residual between the calculated and real observations,

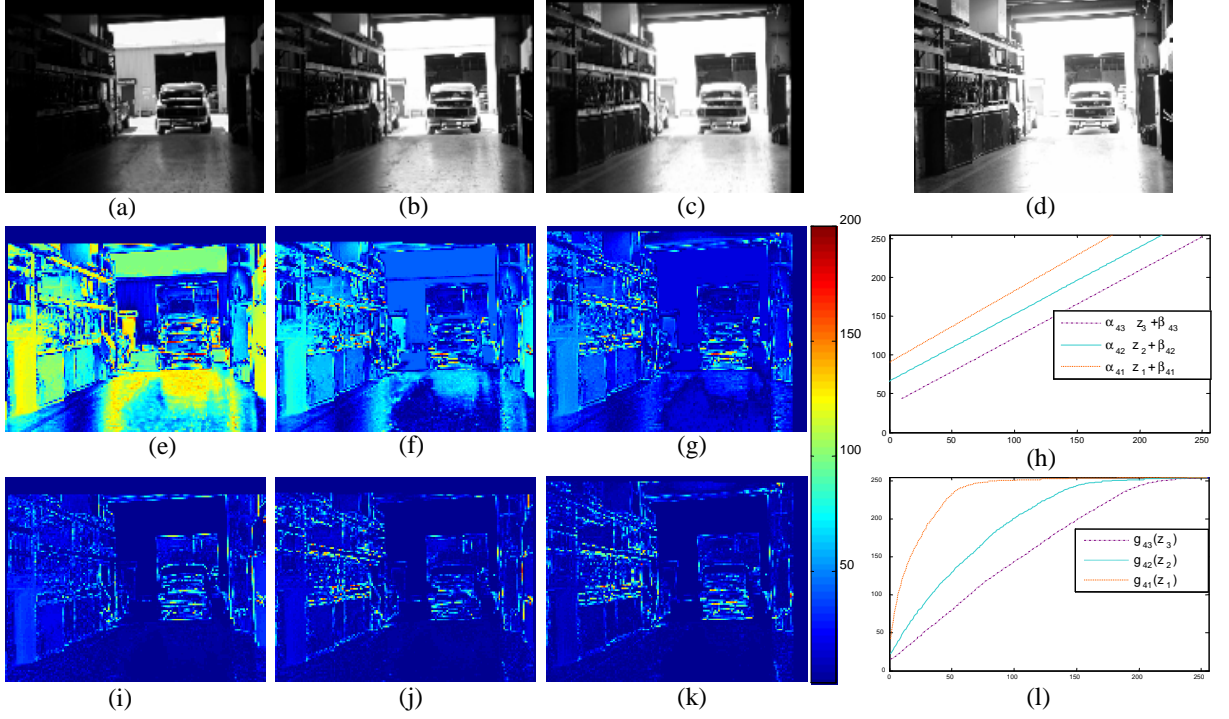


Fig. 18.: Comparison of affine and nonlinear photometric conversion. (a)-(d) are the images captured with different exposure rates. All camera parameters other than the exposure rates are fixed. The images are geometrically registered. The relative exposure rates are as follows: (a) Image \mathbf{z}_1 with exposure rate $1/16$. (b) Image \mathbf{z}_2 with exposure rate $1/4$. (c) Image \mathbf{z}_3 with exposure rate $1/2$. (d) Image \mathbf{z}_4 with exposure rate 1 . Image \mathbf{z}_4 is set as the reference image and other images are photometrically registered to it. The residuals and the registration parameters are shown. (e) Residual between \mathbf{z}_4 and photometrically aligned \mathbf{z}_1 using the affine model. (f) Residual between \mathbf{z}_4 and photometrically aligned \mathbf{z}_2 using the affine model. (g) Residual between \mathbf{z}_4 and photometrically aligned \mathbf{z}_3 using the affine model. (h) The photometric mappings for (e)-(g). (i) Residual between \mathbf{z}_4 and photometrically aligned \mathbf{z}_1 using the nonlinear model. (j) Residual between \mathbf{z}_4 and photometrically aligned \mathbf{z}_2 using the nonlinear model. (k) Residual between \mathbf{z}_4 and photometrically aligned \mathbf{z}_3 using the nonlinear model. (l) The photometric mappings for (i)-(k).

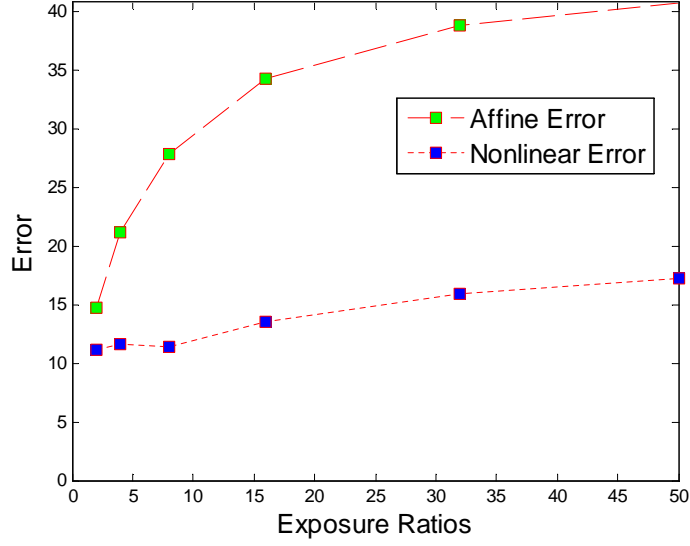


Fig. 19.: Root mean square error (RMSE) values of photometrically registered images with relative exposure rates of 2, 4, 8, 16, 32, 50. The RMSE values (green points) for affine mappings are 14.8, 21.1, 27.8, 34.3, 38.7, 40.7. The RMSE values (blue points) for nonlinear mappings are 11.2, 11.6, 11.4, 13.5, 15.9, 17.2.

and projects the residual back onto the initial estimate. When the imaging model is not accurate or registration parameters are not estimated correctly, the algorithm would fail. In this section, we conclude that nonlinear photometric models should be a part of SR algorithms when there is a possibility of photometric diversity among input images.

C. SR under Photometric Diversity

When all input images are not photometrically identical, there are two possible ways to enhance a reference image: (i) spatial resolution enhancement and (ii) spatial resolution and dynamic range enhancement. In (i), only spatial resolution of the reference image is improved. This requires photometric mapping of all input data to the reference image. In (ii), both spatial resolution and dynamic range of the reference image are improved. This can be considered as a combination of high-dynamic-range imaging and super-resolution image restoration.

1. Spatial Resolution Enhancement

In spatial resolution enhancement, all input images are converted to the tonal range of reference image. After photometric registration, a traditional SR reconstruction algorithm can be applied. However, this is not a straightforward process when the intensity mapping is nonlinear. Refer to Figure 20 that shows various intensity mapping functions (IMFs). Suppose that \mathbf{z}_1 is the reference image to be enhanced. Input image \mathbf{z}_2 is aimed to be photometrically converted onto \mathbf{z}_1 in all cases. There are four cases in Figure 20:

- In case (a), the input image \mathbf{z}_2 has the same photometric range with the reference image; so there is no photometric registration necessary.

- In case (b), the IMF is nonlinear; however, there is no saturation. Therefore, the intensities of \mathbf{z}_2 can be mapped onto the range of \mathbf{z}_1 using the IMF without loss of information.

- In case (c), there is bright saturation in \mathbf{z}_2 . The IMF is not a one-to-one mapping. The problematic region is where the slope of the IMF is zero or close to zero. For saturated regions, there is no information in \mathbf{z}_2 corresponding to \mathbf{z}_1 . Therefore, perfect photometric mapping from \mathbf{z}_2 to \mathbf{z}_1 is not possible. When additive sensor noise and quantization are considered, small-slope (referring to the slope of the IMF) regions would also be problematic in addition to the zero-slope (saturation) regions. In these regions, noise and quantization error in \mathbf{z}_2 would be amplified when mapped onto \mathbf{z}_1 , and reconstruction would be affected negatively.

- In case (d), there are regions of small slope and large slope. Large-slope regions are not an issue because mapping from \mathbf{z}_2 to \mathbf{z}_1 would not create any problem. The problem is still with the small-slope regions (dark saturation regions in \mathbf{z}_2), where quantization and noise floor are effective.

One solution to the saturation and noise amplification problems is to use a certainty function associated with each image. The certainty function should weight

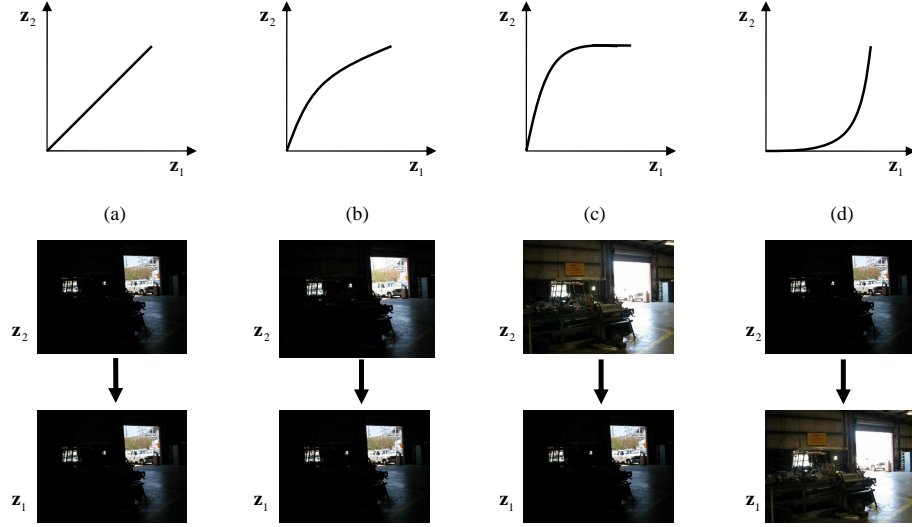


Fig. 20.: Various photometric conversion scenarios. First row illustrates possible photometric conversion functions. Second and third rows show example images with such photometric conversion.

the contribution of each pixel in reconstruction based on the reliability of conversion. If a pixel is saturated or close to saturation, then the certainty function should be close to zero. If a pixel is from a reliable region, then the certainty function should be close to one. The issue of designing a certainty function has been investigated in HDR research. In [38], the certainty function is defined according to the derivative of the CRF. The motivation is that for pixels corresponding to low-slope regions of the CRF, the reliability should also be low. In [56], the certainty function includes variances of the additive noise and quantization errors in addition to the derivative of the CRF. In [34], a fixed hat function is used. According to it, the mid-range pixels have high reliability, while low-end and high-end pixels have low reliability.

We now put these ideas in SR reconstruction. Let \mathbf{x} be the (unknown) high-resolution version of a reference image \mathbf{z}_r , and define $g_{ri}(\mathbf{z}_i)$ as the IMF that takes \mathbf{z}_i and converts it to the photometric range of \mathbf{z}_r , (therefore, \mathbf{x}). Referring to equation (5), $g_{ri}(\mathbf{z}_i)$ includes the CRF $f(\cdot)$, and gain α_{ri} and offset β_{ri} parameters:

$$g_{ri}(\mathbf{z}_i) \equiv f(\alpha_{ri}f^{-1}(\mathbf{z}_i) + \beta_{ri}). \quad (3.6)$$

We also need to model spatial transformations of the imaging process. Define \mathbf{H}_i as the linear mapping that takes a high-resolution image and produces a low-resolution image. \mathbf{H}_i includes motion (of the camera or the objects in the scene), blur (caused by the point spread function of the sensor elements and the optical system), and downsampling. (Details of \mathbf{H}_i modeling can be found in the special issue of the IEEE Signal Processing Magazine [5] and the references therein.)

When \mathbf{H}_i is applied on \mathbf{x} , it should produce the photometrically converted i th observation, $g_{ri}(\mathbf{z}_i)$. That is, we need to find \mathbf{x} that produces $g_{ri}(\mathbf{z}_i)$ when \mathbf{H}_i is applied to it, for all i . The least squares solution to this problem would minimize the following cost function:

$$C(\mathbf{x}) = \sum_i \|g_{ri}(\mathbf{z}_i) - \mathbf{H}_i \mathbf{x}\|^2. \quad (3.7)$$

As explained earlier, the problem associated with the saturation of the IMF can be solved using a certainty function, $w(\mathbf{z}_i)$. We formulate our equations using a generic function $w(\mathbf{z}_i)$. Our specific choice will be given in the experimental results section. We now define a diagonal matrix \mathbf{W}_i whose diagonal is $w(\mathbf{z}_i)$, and incorporate this matrix into equation (3.7) to find the weighted least squares solution. The new cost function is

$$C(\mathbf{x}) = \frac{1}{2} \sum_i (g_{ri}(\mathbf{z}_i) - \mathbf{H}_i \mathbf{x})^T \mathbf{W}_i (g_{ri}(\mathbf{z}_i) - \mathbf{H}_i \mathbf{x}). \quad (3.8)$$

Since dimensions of the matrices are large, we want to avoid matrix inversion and apply the gradient descent technique to find \mathbf{x} that minimizes this cost function. Starting with an initial estimate $\mathbf{x}^{(0)}$, each iteration updates $\mathbf{x}^{(0)}$ in the direction of the negative gradient of $C(\mathbf{x})$:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \gamma \sum_i \mathbf{H}_i^T \mathbf{W}_i (g_{ri}(\mathbf{z}_i) - \mathbf{H}_i \mathbf{x}^{(k)}), \quad (3.9)$$

where γ is the step size at the k th iteration. Defining Φ as the negative gradient of

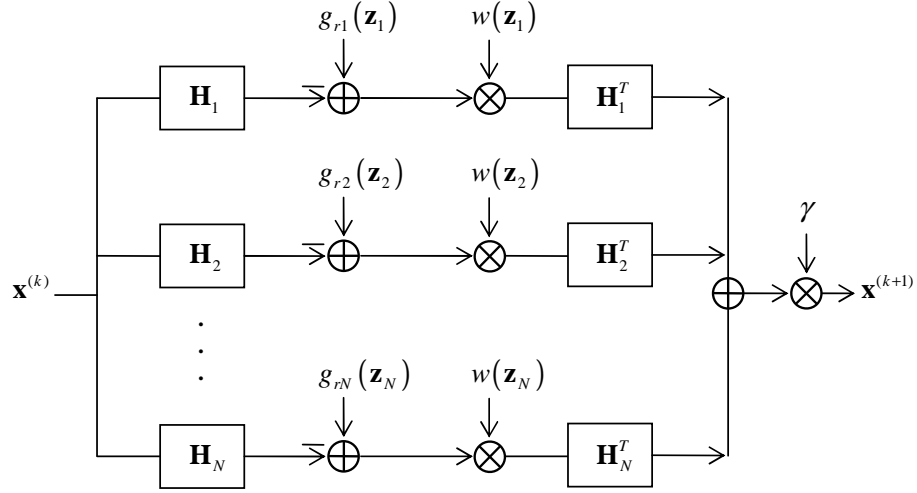


Fig. 21.: Spatial resolution enhancement framework using IMF. $g_{ri}(\cdot)$ is the IMF that converts input image \mathbf{z}_i to the photometric range of reference image. \mathbf{H}_i applies spatial transformations, consisting of geometric warping, blurring, and downsampling. Similarly, \mathbf{H}_i^T applies upsampling with zero insertion, blurring, and back-warping. γ is the step size of the update; it is computed at each iteration.

$C(\mathbf{x})$, the exact line search that minimizes $C(\mathbf{x}^{(k)} + \gamma\Phi)$ yields the step size:

$$\gamma = \frac{\Phi^T \Phi}{\Phi^T \left[\sum_i \mathbf{H}_i^T \mathbf{W}_i \mathbf{H}_i \right] \Phi}, \quad (3.10)$$

with

$$\Phi = \sum_i \mathbf{H}_i^T \mathbf{W}_i (g_{ri}(\mathbf{z}_i) - \mathbf{H}_i \mathbf{x}^{(k)}). \quad (3.11)$$

An iteration of the algorithm is illustrated in Fig. 21, and the pseudocode is given in Table 1. Note that, in implementation, it is not necessary to construct matrices or vectors to follow the steps of the algorithm. Application of \mathbf{H}_i can be implemented as warping an image geometrically, convolving with the point spread function (PSF), and downsampling. Similarly, \mathbf{H}_i^T can be implemented as upsampling with zero insertion, convolving with a flipped PSF, and back-warping [56]. The step size γ can be obtained using the same principles.

Table 3.: Pseudocode of the spatial enhancement algorithm

1. Requirements:
 - Set or estimate the point spread function (PSF) of the camera: \mathbf{h}
 - Set the resolution enhancement factor: F
 - Set the number of iterations: K
2. Initialization:
 - Choose the reference image \mathbf{z}_r
 - Interpolate \mathbf{z}_r by the enhancement factor F to obtain $\mathbf{x}^{(0)}$
3. Parameter estimation:
 - Estimate the spatial registration parameters between \mathbf{z}_r and \mathbf{z}_i , $i = 1, \dots, N$
 - Estimate the IMFs, $g_{ri}(\mathbf{z}_i)$, between \mathbf{z}_r and \mathbf{z}_i , $i = 1, \dots, N$
4. Iterations:
 - For $k = 0$ to $K - 1$
 - Create a zero filled initial image Ψ with the same size as $\mathbf{x}^{(0)}$
 - For $i = 1$ to N
 - Find $\mathbf{H}_i \mathbf{x}^{(k)}$ with the following steps:
 - Convolve $\mathbf{x}^{(k)}$ with the PSF \mathbf{h}
 - Warp and downsample the convolved image onto the input \mathbf{z}_i
 - Find the residual $g_{ri}(\mathbf{z}_i) - \mathbf{H}_i \mathbf{x}^{(k)}$
 - Find the weight image $w(\mathbf{z}_i)$ and multiply it pixel-by-pixel with the residual $g_{ri}(\mathbf{z}_i) - \mathbf{H}_i \mathbf{x}^{(k)}$
 - Obtain $\mathbf{H}_i^T \mathbf{W}_i (g_{ri}(\mathbf{z}_i) - \mathbf{H}_i \mathbf{x}^{(k)})$ with the following steps:
 - Upsample the weighted residual by the factor F with zero insertion
 - Convolve the result with the flipped PSF \mathbf{h}
 - Warp the result to the coordinates of $\mathbf{x}^{(k)}$
 - Update Ψ : $\Psi \leftarrow \Psi + \mathbf{H}_i^T \mathbf{W}_i (g_{ri}(\mathbf{z}_i) - \mathbf{H}_i \mathbf{x}^{(k)})$
 - Calculate γ
 - Update the current estimate: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \gamma \Psi$

2. Spatial Resolution and Dynamic Range Enhancement

Here, the goal is to produce a high-resolution and high-dynamic range image. One option is to obtain the high-resolution version of each input image using the algorithm given in Table 1, and then apply HDR image construction to these high-resolution images. A second option is to derive the high-resolution irradiance, \mathbf{q} , directly.

This requires formulating the image acquisition from the unknown high-resolution irradiance \mathbf{q} to each observation \mathbf{z}_i . Adding the spatial processes (geometric warping, blurring with the PSF, and downsampling) to equation (4), the imaging process can be formulated as

$$\mathbf{z}_i = f(a_i \mathbf{H}_i \mathbf{q} + b_i), \quad (3.12)$$

where \mathbf{H}_i is the linear mapping (including warping, blurring, and downsampling operations) from a high-spatial-resolution irradiance signal to a low-spatial-resolution irradiance signal. $f(\cdot)$, a_i , and b_i are the CRF, gain, and offset terms as in (4).

This time the weighted least squares estimate of \mathbf{q} minimizes the following cost function:

$$C(\mathbf{q}) = \frac{1}{2} \sum_i \left(\frac{f^{-1}(\mathbf{z}_i) - b_i}{a_i} - \mathbf{H}_i \mathbf{q} \right)^T \mathbf{W}_i \left(\frac{f^{-1}(\mathbf{z}_i) - b_i}{a_i} - \mathbf{H}_i \mathbf{q} \right) \quad (3.13)$$

This cost function is basically analogous to the cost function in equation (8). Starting with an initial estimate for \mathbf{q} , the rest of algorithms work similar to the one in Table 1. The only difference is that intensity-to-intensity mapping $g_{ri}(\cdot)$ in (8) is replaced with intensity-to-irradiance mapping $\frac{f^{-1}(\cdot) - b_i}{a_i}$. Unlike the intensity-to-intensity mapping, intensity-to-irradiance mapping requires explicit estimation of the CRF, gain and offset parameters.

We write the iterative step to estimate \mathbf{q} :

$$\mathbf{q}^{(k+1)} = \mathbf{q}^{(k)} + \gamma \sum_i \mathbf{H}_i^T \mathbf{W}_i \left(\frac{f^{-1}(\mathbf{z}_i) - b_i}{a_i} - \mathbf{H}_i \mathbf{q}^{(k)} \right) \quad (3.14)$$

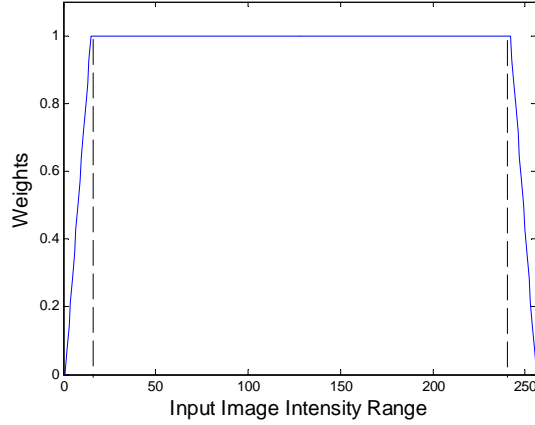


Fig. 22.: Piecewise linear certainty function used in the experiments. The intensity break-points in the figure are 15 and 240.

where γ is the step size. It is obtained similar to the one in equation (9). The details of this approach is trivial given the derivations in the previous section; therefore, we leave it to the reader.

In [56], we also investigated this joint spatial and dynamic range enhancement idea. The approach in [56] is similar to the one (irradiance-domain solution) given in this section. As we mentioned in Section I, in [56], we applied Taylor series expansion to the inverse of the CRF to end up with a specific certainty function. The algorithm requires estimation of the CRF and the variances of noise and quantization error. It also includes a spatial regularization term in the reconstruction. The derivation in this section can be considered as a generalization of the solution given in [56]; here, the certainty function is not specified. In practice, the methods in [56] and the irradiance-domain solution of this section work similarly with the proper selection of certainty functions.

Note that this approach estimates the irradiance \mathbf{q} , which needs to be compressed in dynamic range to display on limited dynamic range displays. Displaying HDR images on limited dynamic range displays is an active research area [64].

3. Certainty Function

As we have discussed in Section III-A, the information coming from low-end and high-end of the intensity range is not reliable due to noise, quantization, and saturation. If used, these unreliable pixels would degrade the restoration. In [34], a generalized hat function is proposed to reduce the effect of unreliable pixels. We use a piecewise linear certainty function in our experiments. The certainty function is shown in Figure 22. The intensity breakpoints in the certainty function are 15 and 240, and they were determined by trial-and-error.

Figure 23 shows an example to demonstrate the reliability of pixels and the effect of the certainty function. The first row in the figure shows photometric conversion from an over-exposed image to an under-exposed image. This is the scenario in Figure 3(c). Figure 23(a) is the reference image, and Figure 23(b) is the geometrically warped input image which we want to map onto the reference image tonally. Figure 23(c) shows the residual between the input and the reference image without photometric registration. Figure 23(d) shows the residual after the application of IMF to the input image. Clearly, saturated pixels are not handled well: residuals for these pixels are large. The weights for each pixel in the image are calculated with application of the certainty function on the input image; they are shown in Figure 23(e). Examining Figure 23(d) and (e), it can be seen that the weights for unreliable saturated pixels are low, as expected. Figure 23(f) shows the final residual after the application of the weight image Figure 23(e) on the residual image Figure 23(d).

The second row in Figure 23 shows an example of tonal conversion from an under-exposed image to an over-exposed image. This is the scenario in Figure 3(d). Figure 23(g) is the reference image and Figure 23(h) is the geometrically warped input image. Here, photometric transformation can be performed without any problem for high-end of intensity range. The problem is the low-end, dark saturation regions in the input image. Figure 23(j) shows the residual after tonal conversion. Figure 23(k)

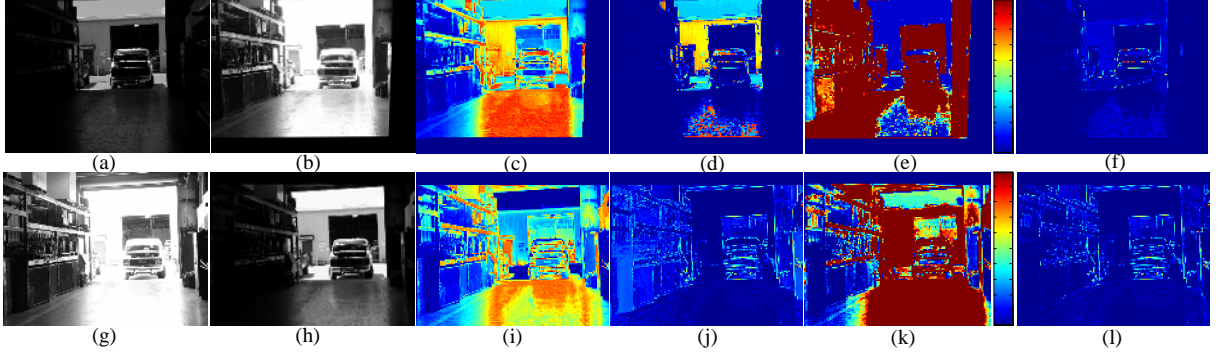


Fig. 23.: Weighting function effect on residuals. First row performs conversion onto low exposure reference while second row has a reference with high exposure. (a) Reference image. (b) Geometrically warped input image. (c) Residual image without tonal conversion. (d) Residual image using nonlinear tonal conversion. (e) Certainty image using hat function as weighting and image (b) as input. (f) Weighted residual obtained multiplying residual image in (d) by certainty image in (e). (g) Reference image. (h) Geometrically warped input image. (i) Residual image without tonal conversion. (j) Residual image using nonlinear tonal conversion. (k) Certainty image using hat function as weighting and image (g) as input. (l) Weighted residual obtained multiplying residual image in (j) by certainty image in (k).

is the certainty image. As seen, the unreliable dark saturation regions are having low weights. Figure 23(f) and (l) show the weighted residuals obtained by multiplying the residual images with the corresponding certainty images. In the weighted residual image, large residual values (that would degrade SR reconstruction) are eliminated or reduced significantly.

D. Geometric and Photometric Registration

SR requires accurate geometric and photometric registration. If the actual CRF and the exposure rates are unknown, the images must be geometrically registered before these parameters can be estimated. On the other hand, geometric registration is problematic when images are not photometrically registered. There are three possible approaches to this problem:

- (A1) Images are first geometrically registered using an algorithm that is insensitive to photometric changes. This is followed by photometric registration.

- (A2) Images are first photometrically registered using an algorithm that is insensitive to geometric misalignments. This is followed by geometric registration.
- (A3) Geometric and photometric registration parameters are estimated jointly.

There are few algorithms that can be utilized for these approaches. In [65], an exposure-insensitive motion estimation algorithm based on the Lucas-Kanade technique is proposed to estimate motion vectors at each pixel. Although this method can be used to estimate large and dense motion field, it has the downside that it requires pre-knowledge of the CRF. Another exposure-insensitive algorithm is proposed in [66]. It is based on bit-matching on binary images. Although it does not require knowing CRF in advance, the algorithm is limited to global translational motion. In [60], an IMF estimation algorithm that does not require geometric registration is proposed. It is based on the idea that histogram specification gives the intensity mapping between two images when there is no saturation or significant geometric misalignment. And finally in [67], a joint geometric and photometric registration algorithm was proposed. There, the problem is formulated as a global parameter estimation, where the parameters jointly represent geometric transformation, exposure rate, and CRF. Two potential problems associated with this approach are (1) getting stuck at a local minima and (2) limitation of using parametric CRF.

We take approach (A1) in our experiments. This is also the approach in [56]. [41, 55] takes the same approach except for the photometric model. For geometric registration, we use a feature-based algorithm, which requires robust exposure-insensitive feature extraction and matching. In our experiments, feature points are first extracted using the Harris corner detector [43]. Although the Harris corner detector is not invariant to illumination changes in general, it worked well in our experiments. These feature points are matched using normalized cross correlation, which is insensitive to contrast changes. The RANSAC method [44] is then used to eliminate the outliers and estimate the homographies. After geometric registration comes photometric reg-

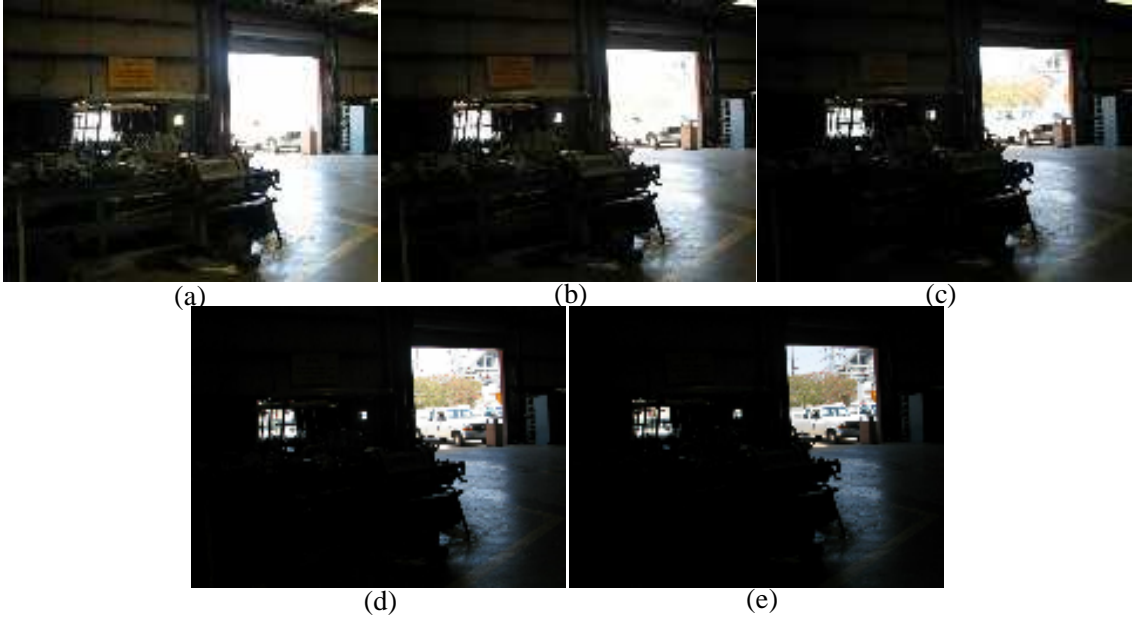


Fig. 24.: Five images of “Facility I” data set that includes 22 images are displayed here. Exposure durations of (a)-(e) are $\frac{1}{25}$, $\frac{1}{50}$, $\frac{1}{100}$, $\frac{1}{200}$, and $\frac{1}{400}$ seconds, respectively.

istration. There are various methods available in the literature to estimate IMF and CRF as we discussed earlier. In our experiments, we use [62] to estimate IMF, CRF and the exposure rate.

E. Experiments and Results

We conducted experiments to demonstrate the proposed SR algorithms. (A MATLAB toolbox can be downloaded from [68].) We captured two data sets with a hand-held digital camera. These data sets are shown in Figure 24 and 25. The resolution enhancement factor is four and the number of iterations was set to two in all experiments. The PSF is taken as a Gaussian window of size $[7 \times 7]$ and of variance 1.7. The results are shown in Figure 26 and Figure 27. For the spatial-only enhancement approach, we did experiments when the reference is chosen as an over-exposed image and also when it is chosen as an under-exposed image to show the robustness of the algorithm. For the irradiance-domain spatial and dynamic range enhancement approach, we created an initial estimate by applying a standard HDR image construction algorithm [34]. The initial estimate is then updated iteratively.

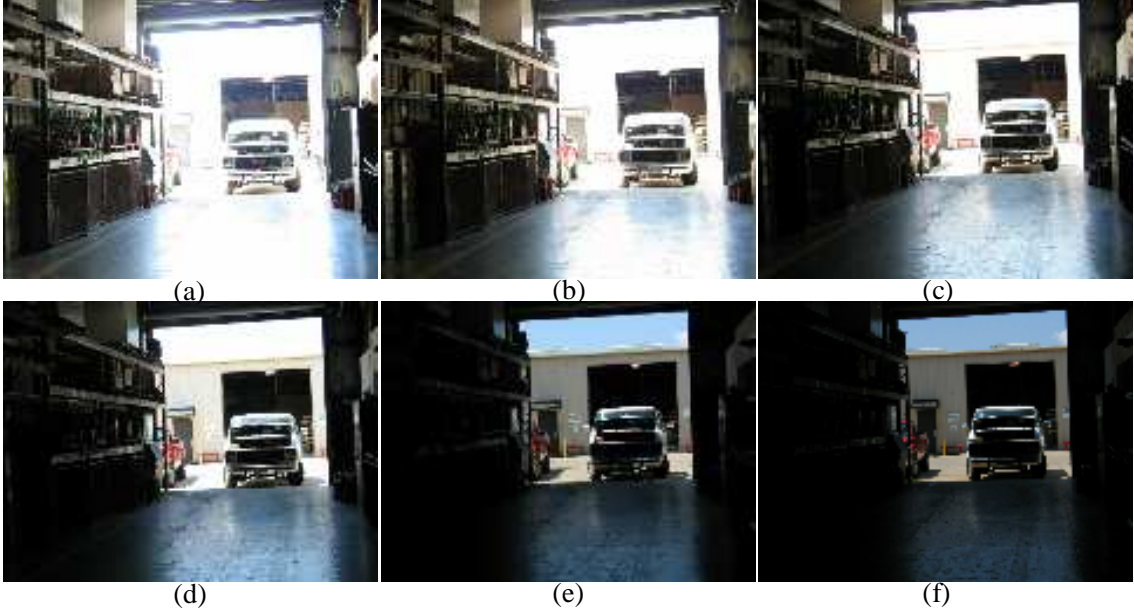


Fig. 25.: Six images of “Facility II” data set that includes 31 images are displayed here. Exposure durations of (a)-(e) are $\frac{1}{25}$, $\frac{1}{50}$, $\frac{1}{100}$, $\frac{1}{200}$, $\frac{1}{400}$, and $\frac{1}{800}$ seconds, respectively.

For comparison purposes, we provided cropped regions of results and transformed input images in Figure 26. Figure 26(a)-(e) are the bilinearly interpolated observations with different exposure rates chosen from the 22 input images. Each observation contains a different portion of the existing tonal range. Figure 26(f) is the SR result obtained when the over-exposed image, Figure 26(a), is chosen as the reference image. Comparison of figures 26(a) and 26(f) shows the improvement in resolution. Similarly, Figure 26(g) is obtained when the under-exposed image, 26(e), is set as the reference image. The resolution enhancement is again clear. Figure 26(h) is the result of the resolution and dynamic range enhancement algorithm. Notice that both spatial resolution and dynamic range are improved. (Note that the result of an HDR algorithm would have higher dynamic range than a standard display or printing device. Therefore, an HDR image should be compressed in range to output in such LDR devices. There are complex display algorithms [64]. In this paper, we used a simple gamma correction to display the result. The gamma correction parameter is

0.5.) Figure 2 shows the results for the second data set. The discussion is parallel to the discussion of first data set, therefore, is excluded for conciseness.

Finally, we wanted to test of the effect of the weighting function in SR reconstruction. Figure 28(a) and Figure 28(b) aim to increase both spatial resolution and dynamic range, and differ only in their weighting function. Figure 28(a) shows SR result using a unity weight. Notice the loss of information in details and color. Color artifacts occur in this case as the saturated pixels are not handled properly. Strips on the warehouse can hardly be observed and the sky color is washed out due to fusion with saturated residuals. There is also contouring artifact. Figure 28(b) shows the result using the proposed hat function for weighting. Texture and colors are preserved compared to Figure 28(a).

F. Conclusions

In this paper, we showed how to do SR when the photometric characteristics of the input images are not identical. We showed two possible approaches, one of them enhancing spatial resolution only, and the other enhancing spatial resolution and the dynamic range jointly. We demonstrated that nonlinear photometric modeling should be preferred to affine photometric modeling. We also discussed that an appropriate weighting function is necessary to handle saturation. Other SR reconstruction techniques can be modified and applied as long as an appropriate photometric registration is included.

Although geometric registration is a complicated task for differently exposed images, we achieved a success using feature-based registration without the need of CRF. A sequential approach is useful: Similarly exposed images would produce similar features; therefore, correct geometric registration would be easier to estimate among similarly exposed images compared to images with large exposure differences. Therefore, one may find the homographies among similarly exposed images, and then combine the homographies to find the homography between any two image. In our



Fig. 26.: Cropped regions of observation and SR results. (a)-(e) Input images. (f) SR when (a) is the reference image. (g) SR when (e) is the reference image. (h) SR using the technique presented in Section 2.



Fig. 27.: Cropped regions of observation and SR results. (a)-(d) Input images. (e) SR when (d) is the reference image. (f) SR when (a) is the reference image. (g) SR using the technique presented in Section 2.



Fig. 28.: Comparison of weighting function during spatial resolution enhancement. Lowest exposed image is chosen as reference. (a) SR reconstruction using identity weight, (b) SR reconstruction using a hat function as weight.

experiments, we chose a reference image from the middle of the tonal range. Homographies of input images are found with respect to this reference image. The homography between any two of the input images can then be found by multiplying the corresponding homography matrices. Photometric registration is trivial after the geometric registration.

In this paper, we considered static scenes only. Estimation of motion parameters for non-static scenes is left as a future work. Also, we only considered global photometric changes. In experiments, we only considered the exposure rate change which acts globally on image. In general, a dense photometric model is necessary to handle local photometric changes. That is, accurate geometric and photometric registration of photometrically different images is certainly a problem requiring further research.

CHAPTER IV

RESTORATION OF BAYER-SAMPLED IMAGE SEQUENCES

Spatial resolution of digital images are limited due to optical/sensor blurring and sensor site density. In single-chip digital cameras, the resolution is further degraded because such devices use a color filter array to capture only one spectral component at a pixel location. The process of estimating the missing two color values at each pixel location is known as demosaicking. Demosaicking methods usually exploit the correlation among color channels. When there are multiple images, it is possible not only to have better estimates of the missing color values but also to improve the spatial resolution further (using super-resolution reconstruction). In this paper, we propose a multi-frame spatial resolution enhancement algorithm based on the projections onto convex sets (POCS) technique.

A. Introduction

In a digital camera, there is a rectangular grid of electron-collection sites laid over a silicon wafer to measure the amount of light energy reaching each of them. When photons strike these sensor sites, electron-hole pairs are generated; and the electrons generated at each site are collected over a certain period of time. The numbers of electrons are eventually converted to pixel values.

To produce color pictures, this image acquisition process is modified in various ways. The first method is to use beam-splitters to split the light into several optical paths and use different spectral filters on each path to capture different spectral components. This method requires precise alignment of images from each color channel. Another possibility is to capture multiple pictures with a single sensor; each time a different color filter placed in front of the whole sensor. The alternative and less-expensive method is to use a mosaic of color filters to capture only one spectral component at each sensor site. The color samples obtained with such a color filter array (CFA) must then be interpolated to estimate the missing samples. Because of

the mosaic pattern of the color samples, this CFA interpolation problem is also referred to as “demosaicking” in the literature [69]. Although the last method causes a loss of spatial resolution in every color channel, it is usually the preferred one because of its simplicity and lower cost.

The most commonly used CFA pattern is the “Bayer” pattern, which is shown in Figure 29. In a Bayer pattern, green samples are obtained on a quincunx lattice, and red and blue samples are obtained on rectangular lattices. The green channel is more densely sampled than the red and blue channels because the spectral response of a green filter is similar to the human eye’s luminance frequency response.

If the measured image is divided by measured color into three separate images, this problem looks like a typical image interpolation problem. Therefore, standard image interpolation techniques can be applied to each channel separately. Bilinear and bicubic interpolations are common image interpolation techniques that produce good results when applied to gray-scale images. However, when they are used for the demosaicking problem, the resulting image shows many visible artifacts. This motivates the need to find a specialized algorithm for demosaicking problem. There is a variety of demosaicking algorithms in the literature, including edge-directed [70, 2], constant-hue-based [71, 72, 73, 74], Laplacian as correction terms [75], alias canceling [76], projections onto convex sets based [77], Bayesian [78], artificial neural networks based [79], and homogeneity-directed [80] interpolation methods. Two important ideas in these algorithms are avoiding interpolation across edges and exploiting the correlation among the color channels. These two ideas are put into work in various ways.

When there are multiple images, spatial resolution can be improved beyond the physical limits of a sensor chip. This multi-frame resolution enhancement is known as super-resolution reconstruction. Although there have been a significant amount of work on super-resolution image reconstruction, there are only few recent efforts

modeling CFA sampling and addressing issues related to image reconstruction from CFA-sampled data.

In [3], data fidelity and regularization terms are combined to produce high-resolution images. The data fidelity term is based on a cost function that consists of the the sum of residual differences between actual observations and high-resolution image projected onto observations (*simulated* observations). Regularization functions are added to this cost function to eliminate color artifacts and preserve edge structures. These additional constraints are defined as luminance, chrominance, and orientation regularization in [3]. A similar algorithm is presented in [81].

Among different demosaicking approaches, the POCS-based demosaicking algorithm has one of the best performance [77], [69]. One advantage of the POCS framework is that new constraints can be easily incorporated into a POCS-based algorithm. Here, we present POCS-based algorithms for multi-frame resolution enhancement of Bayer CFA sampled images. We investigate two approaches: *demosaicking and super-resolution* (DSR) and *only super-resolution* (OSR). In DSR approach, we separate the restoration process into two parts. The first part is multi-frame demosaicking, where missing color samples are estimated. The second part is super-resolution reconstruction, where sub-pixel level resolution is achieved. In OSR approach, super-resolution image is reconstructed without demosaicking, using Bayer pattern masks only for computing the residual of each color channel.

In Section B, we present an imaging model that includes CFA sampling. We introduce our POCS-based multi-frame demosaicking algorithm in Section C. This Section includes three constraint sets and corresponding projection operators for multi-frame demosaicking. In Section D, we explain how to achieve sub-pixel resolution. Experimental results and discussions are provided in Section E.

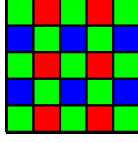


Fig. 29.: Single-chip digital cameras have color filter arrays with specific patterns. The most commonly used pattern is the “Bayer” pattern, which is illustrated in this picture.

B. Imaging Model

We use an image acquisition model that includes CFA sampling in addition to motion and blurring. The model has two steps: the first step models conversion of a high-resolution full-color image into a low-resolution full-color image. The second step models color filter sampling of full-color low-resolution image.

Let \mathbf{x}_S a color channel of a high-resolution image, where a channel can be red (\mathbf{x}_R), green (\mathbf{x}_G), and blue (\mathbf{x}_B). The i th observation, $\mathbf{y}_S^{(i)}$, is obtained from this high-resolution image through spatial warping, blurring, and downsampling operations:

$$\mathbf{y}_S^{(i)} = DCW^{(i)}\mathbf{x}_S, \quad \text{for } S = R, G, B, \text{ and } i = 1, 2, \dots, K, \quad (4.1)$$

where K is the number of input images, $W^{(i)}$ is the warping operation (to account for the relative motion between observations), C is the convolution operation (to account for the point spread function of the camera), and D is the downsampling operation (to account for the spatial sampling of the sensor).

The full-color image ($\mathbf{y}_R^{(i)}, \mathbf{y}_G^{(i)}, \mathbf{y}_B^{(i)}$) is then converted to a mosaicked observation $\mathbf{z}^{(i)}$ according to a CFA sampling pattern:

$$\mathbf{z}^{(i)} = \sum_{S=R,G,B} M_S \mathbf{y}_S^{(i)} \quad (4.2)$$

where M_S takes only one of the color samples at a pixel according to the pattern. For example, at red pixel location, $[M_R, M_G, M_B]$ is $[1, 0, 0]$.

We investigate two approaches to obtain \mathbf{x}_S . In DSR approach, each full-color low-resolution image $\mathbf{y}_S^{(i)}$ is estimated from the multiple observations $\mathbf{z}^{(i)}$, and then

super-resolution reconstruction is applied to these estimates $\mathbf{y}_S^{(i)}$ to obtain \mathbf{x}_S . In the OSR approach, \mathbf{x}_S is obtained directly from $\mathbf{z}^{(i)}$ without estimating $\mathbf{y}_S^{(i)}$.

C. Multi-Frame Demosaicking

In this section, we introduce multi-frame demosaicking, where multiple images are used to estimate missing samples in a Bayer-sampled color image. The process is based on the POCS technique, where an initial estimate is projected onto convex constraint sets iteratively to reach a solution that is consistent with all constraints about the solution [82].

This algorithm is based on the demosaicking approach presented in [77]. In [77], we define two constraint sets. The first constraint set stems from inter-channel correlation. It is based on the idea that high-frequency components of red, green, and blue channels should be similar in an image. This turns out to be a very effective constraint set. This constraint set is also used in this paper.

The second constraint set ensures that the reconstructed image is consistent with the observed data of the same image. This constraint set can be improved when there are multiple observations of the same scene. A missing sample (due to the CFA sampling) in an image could have been captured in another image due to the relative motion between observations. It is possible to obtain a better estimate of a missing sample by taking multiple images into account. Therefore, we define a constraint set based on samples coming from multiple images.

In addition to previous constraint sets, we propose a third constraint set in this paper to achieve color consistency. Consistent neighbors of a pixel are determined according to both spatial and intensity closeness. Intensity of a pixel is updated using the weighted average value of its consistent neighbors. These constraint sets are examined in following sections.

1. Detail Constraint Set

Let \mathcal{W}_k be an operator that produces the k th subband of an image. There are four frequency subbands ($k = LL, LH, HL, HH$) corresponding to low-pass filtering and high-pass filtering permutations along horizontal and vertical dimensions. (A brief review of the subband decomposition is provided in the Appendix B.)

The “detail” constraint set (C_d) forces the details (high-frequency components) of the red and blue channels to be similar to the details of the green channel at every pixel location (n_1, n_2) , and is defined as follows:

$$C_d = \left\{ \begin{array}{l} \mathbf{y}_S^{(i)}(n_1, n_2) : \left| \left(\mathcal{W}_k \mathbf{y}_S^{(i)} \right)(n_1, n_2) - \left(\mathcal{W}_k \mathbf{y}_G^{(i)} \right)(n_1, n_2) \right| \leq \mathbf{T}_d(n_1, n_2) \\ \forall (n_1, n_2), \quad \text{for } k = LH, HL, HH \text{ and } S = R, B \end{array} \right\}, \quad (4.3)$$

where $\mathbf{T}_d(n_1, n_2)$ is a positive threshold that quantifies the “closeness” of the detail subbands to each other. For details and further discussion on this constraint set, we refer the reader to [77].

2. Observation Constraint Set

The interpolated color channels must be consistent with the color samples captured by the digital camera for all images. Even if a color sample does not exist in an image as a result of Bayer sampling, that particular sample could have been captured in another frame (due to motion). By warping all captured samples onto the common frame to be demosaicked, we can obtain a good estimate of the missing samples.

Figure 30 illustrates the sampling idea. In the figure, red channels of three Bayer-sampled images are shown. The grid locations with “triangles”, “circles”, and “squares” are the ones that have red samples. We would like to estimate the missing samples in the middle image. The other two images warped onto the middle image. The estimation problem now becomes an interpolation problem from a set of nonuniformly sampled data. After the interpolation, we obtain an “observation” image, $\bar{\mathbf{y}}_S^{(i)}$.

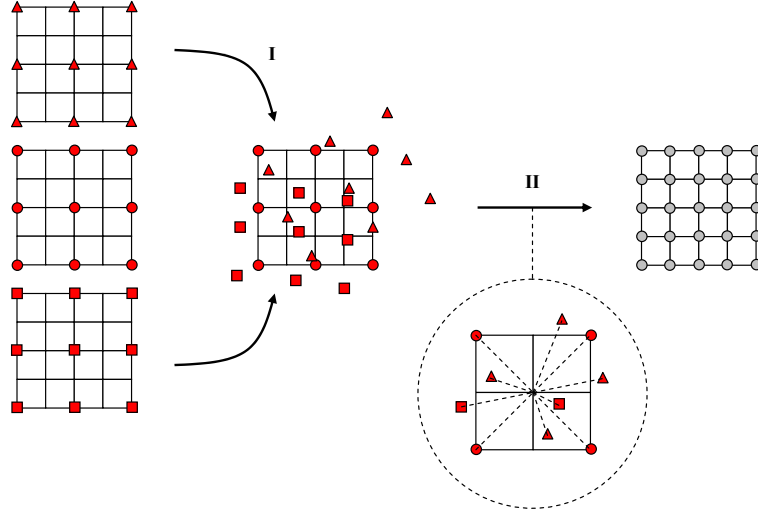


Fig. 30.: I: Input images are warped onto reference frame. II: Weighted average of samples are taken to find values on the reference sampling grid.

The demosaicked samples $\mathbf{y}_S^{(i)}$ should be consistent with this observation image $\bar{\mathbf{y}}_S^{(i)}$.

To make sure that, we can define the “observation” constraint set C_o as follows:

$$C_o = \left\{ \begin{array}{l} \mathbf{y}_S^{(i)}(n_1, n_2) : \left| \mathbf{y}_S^{(i)}(n_1, n_2) - \bar{\mathbf{y}}_S^{(i)}(n_1, n_2) \right| \leq \mathbf{T}_o(n_1, n_2) \\ \forall (n_1, n_2) \text{ and } S = R, G, B \end{array} \right\}, \quad (4.4)$$

where $\mathbf{T}_o(n_1, n_2)$ is a positive threshold to quantify the closeness of the estimated pixels to the averaged data.

The observation image $\bar{\mathbf{y}}_S^{(i)}$ should be interpolated from nonuniformly located samples. There are different approaches to do nonuniform interpolation. One approach is to take a weighted average of all samples within a neighborhood of a pixel in question. For example, if Euclidean distance between a sample and the pixel in question is d , then the weight of the sample could be set to be proportional to $e^{-(d/\sigma)^2}$, where σ is a constant parameter. This requires accurate geometric registration as weight of each sample is determined according to the spatial closeness to the sampling location. In our experiments, we used simple bilinear interpolation to

estimate the missing samples. That is, all frames are warped onto the reference frame using bilinear interpolation. These warped images are then averaged to obtain $\bar{\mathbf{y}}_S^{(i)}$.

One benefit of using this constraint set is noise reduction. When the input data is noisy, constraining the solution to be close to the observation image $\bar{\mathbf{y}}_S^{(i)}$ prevents amplification of noise and color artifacts.

3. Color Consistency Constraint Set

The third constraint set is the color consistency constraint set. It is reasonable to expect pixels with similar green intensities to have similar red and blue intensities within a small spatial neighborhood. Therefore, we define *spatio-intensity neighborhood* of a pixel. Suppose that green channel \mathbf{x}_G of an image is already interpolated and we would like to estimate the red value at a particular pixel (n_1, n_2) . Then, the spatio-intensity neighborhood of the pixel (n_1, n_2) is defined as

$$\mathcal{N}(n_1, n_2) = \left\{ \|(m_1, m_2) - (n_1, n_2)\| \leq \tau_S \text{ and } \left| \mathbf{y}_G^{(i)}(m_1, m_2) - \mathbf{y}_G^{(i)}(n_1, n_2) \right| \leq \tau_I \right\}, \quad (4.5)$$

where τ_S and τ_I are spatial and intensity neighborhood ranges, respectively. Figure 31 illustrates spatio-intensity neighborhood for a one-dimensional signal. In that figure, we would like to determine the spatio-intensity neighbors of the pixel in the middle. The spatial neighborhood is determined by τ_S ; and the intensity neighborhood is determined by τ_I . The yellow region shows the spatio-intensity neighborhood.

The spatio-intensity neighbors of a pixel should have similar color values. One way to measure color similarity is to inspect color differences between red and green channels (and blue and green channels). These differences are expected to be similar within the spatio-intensity neighborhood $\mathcal{N}(n_1, n_2)$. Therefore, the color consistency

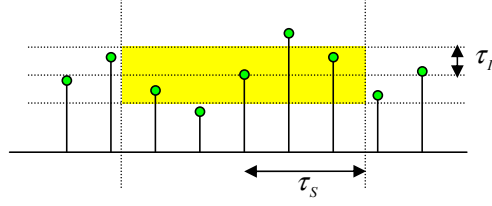


Fig. 31.: Spatio-intensity neighborhood of a pixel is illustrated on a one-dimensional image. The yellow region is the neighborhood of the pixel in the middle.

constraint set can be defined as follows:

$$C_c = \left\{ \begin{array}{l} \mathbf{y}_S^{(i)}(n_1, n_2) : \left| \left(\mathbf{y}_S^{(i)}(n_1, n_2) - \mathbf{y}_G^{(i)}(n_1, n_2) \right) - \overline{\left(\mathbf{y}_S^{(i)}(n_1, n_2) - \mathbf{y}_G^{(i)}(n_1, n_2) \right)} \right| \leq \mathbf{T}_c(n_1, n_2) \\ \forall (n_1, n_2) \text{ and } S = R, B \end{array} \right\}, \quad (4.6)$$

where $\overline{(\cdot)}$ is averaging within the neighborhood $\mathcal{N}(n_1, n_2)$, and $\mathbf{T}_c(n_1, n_2)$ is a positive threshold. We perform the averaging operation $\overline{(\cdot)}$ using the bilateral filter concept [83]. Using a gaussian filter near edges causes interpolation across different regions, which leads to blurring. Bilateral filter was proposed to perform image smoothing without blurring across images. It performs averaging in both spatial domain and intensity range of the neighborhood and can be formulated as follows

$$\overline{I(n_1, n_2)} = \frac{1}{Z} \sum_{m_1, m_2 \in \mathcal{N}(n_1, n_2)} \exp\left(-\frac{d}{2\sigma_S^2}\right) \exp\left(-\frac{(I(m_1, m_2) - I(n_1, n_2))^2}{2\sigma_I^2}\right) I(m_1, m_2), \quad (4.7)$$

where Z is the normalizing constant, and d is the euclidian distance between coordinates, $(m_1 - n_1)^2 + (m_2 - n_2)^2$. Note that when τ_I and τ_S are kept relatively large, σ_I and σ_S become the real governing parameters. Also note that we have defined the spatio-intensity neighborhood within a frame. This can be extended to multiple images using motion vectors. The idea is illustrated in Figure 32.

An issue to be considered with all these constraint sets is the selection of the threshold values $\mathbf{T}_d(n_1, n_2)$, $\mathbf{T}_o(n_1, n_2)$, and $\mathbf{T}_c(n_1, n_2)$. The detail constraint set should be tight (i.e. $\mathbf{T}_d(n_1, n_2)$ should be small) when the inter-channel correlation

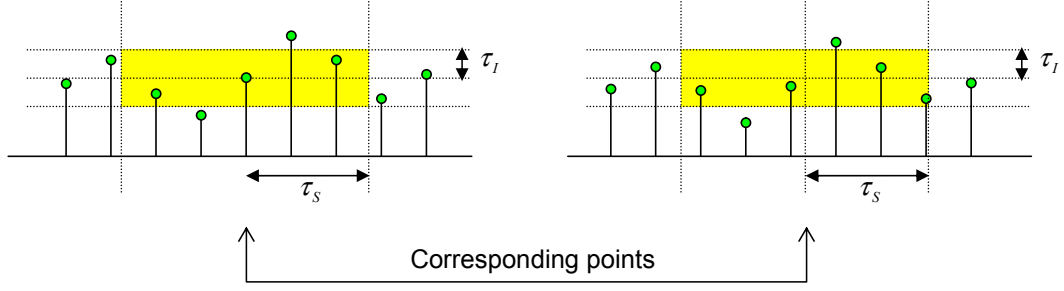


Fig. 32.: The spatio-intensity neighborhood of a pixel can be defined for multiple images. The corresponding point of a pixel is found using motion vectors; using the parameters τ_s and τ_I , the neighborhood (yellow regions) is determined.

is high [77]. When the images are noisy, a strict observation constraint that forces the solution towards the average image $\bar{\mathbf{y}}_S^{(i)}$ would prevent color artifacts. Therefore, noise variance and the number of images used to obtain $\bar{\mathbf{y}}_S^{(i)}$ should be considered during the selection of $\mathbf{T}_o(n_1, n_2)$. If the images were noise-free and the warping process was perfect, then we would keep $\mathbf{T}_o(n_1, n_2)$ very small. The color consistency threshold $\mathbf{T}_c(n_1, n_2)$ determines how much the color difference smoothness is enforced. The selection of the spatio-intensity neighborhood is also critical in this constraint set. When the spatial range τ_s is too large, local spatial neighborhood assumption could be violated. When τ_s is too small, noise might become an issue. On the other hand, the intensity range τ_I should not be too large not to take unrelated pixels into account. In our experiments, we determined optimal parameters by trial-and-error. Selecting the size of the spatio-intensity neighborhood window is a research topic by itself. Spatio-temporal activity of a video plays an important role in parameter selection for color consistency set. We leave the robust and automatic parameter selection issue as a future work.

4. Projection Operations

In this section, we define projection operators for the aforementioned constraint sets. These projection operations will form the basis of joint demosaicking and super-resolution, which will be presented in following sections.

a. Projection onto detail constraint set

We write the projection $P_d \left[\mathbf{y}_S^{(i)} \right]$ of a color channel $\mathbf{y}_S^{(i)}$ onto the detail constraint set C_d as follows [77]:

$$P_d \left[\mathbf{y}_S^{(i)} \right] = \mathcal{U}_{LL} \mathcal{W}_{LL} \mathbf{y}_S^{(i)} + \sum_{k=HL, LH, HH} \mathcal{U}_k \mathcal{W}'_k \mathbf{y}_S^{(i)} \quad (4.8)$$

where

$$\mathcal{W}'_k \mathbf{y}_S^{(i)} = \begin{cases} \mathcal{W}_k \mathbf{y}_G^{(i)} + \mathbf{T}_d, & \text{if } \left(\mathcal{W}_k \mathbf{y}_S^{(i)} - \mathcal{W}_k \mathbf{y}_G^{(i)} \right) > \mathbf{T}_d \\ \mathcal{W}_k \mathbf{y}_S^{(i)}, & \text{if } \left| \mathcal{W}_k \mathbf{y}_S^{(i)} - \mathcal{W}_k \mathbf{y}_G^{(i)} \right| \leq \mathbf{T}_d \\ \mathcal{W}_k \mathbf{y}_G^{(i)} - \mathbf{T}_d, & \text{if } \left(\mathcal{W}_k \mathbf{y}_S^{(i)} - \mathcal{W}_k \mathbf{y}_G^{(i)} \right) < -\mathbf{T}_d \end{cases} \quad (4.9)$$

Detail projection can be implemented efficiently in wavelet domain. \mathcal{W}_k corresponds to decomposing the signal into k th subband, and \mathcal{U}_k is the synthesis filter for the k th subband. Decomposition and synthesis filters should satisfy the perfect reconstruction condition, which is given in Appendix B. Remember that detail projection is applied to red and blue channel, as $S = R, B$ for detail projection.

b. Projection onto observation constraint set

Although detail projection provides consistency of high frequency components of color channels, this forced correlation may result in deviation from actual values. In order to prevent this possible deviation, intermediate result is also projected onto each observations as such

$$P_o \left[\mathbf{y}_S^{(i)} \right] = \begin{cases} \bar{\mathbf{y}}_S^{(i)} + \mathbf{T}_o, & \text{if } \left(\mathbf{y}_S^{(i)} - \bar{\mathbf{y}}_S^{(i)} \right) > \mathbf{T}_o \\ \mathbf{y}_S^{(i)}, & \text{if } \left| \mathbf{y}_S^{(i)} - \bar{\mathbf{y}}_S^{(i)} \right| \leq \mathbf{T}_o \\ \bar{\mathbf{y}}_S^{(i)} - \mathbf{T}_o, & \text{if } \left(\mathbf{y}_S^{(i)} - \bar{\mathbf{y}}_S^{(i)} \right) < -\mathbf{T}_o \end{cases}$$

c. Projection onto color consistency constraint set

Color consistency constraint takes advantage of the fact that correlation of green samples in a small window implies the correlation among red and blue samples. This constraint set provides noise elimination for highly degraded images with noise. Keep-

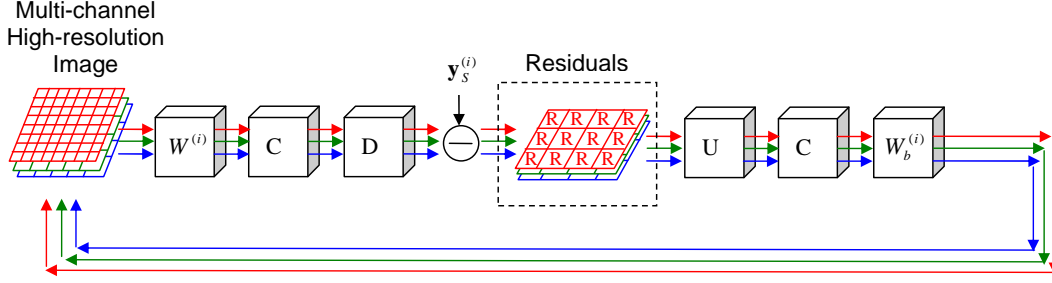


Fig. 33.: DSR reconstruction approach is illustrated for the iterated back-projection algorithm. The algorithm starts with a high-resolution image estimate. Simulated observations are obtained by forward imaging operations. The observations $\mathbf{y}_S^{(i)}$ are obtained using single-frame or multi-frame demosaicking. The full-color residuals are then back-projected to update the current estimate of high-resolution. The notation in the figure is as follows. $W^{(i)}$: Spatial warping onto i th observation, C : Convolution with the PSF, D : Down-sampling by the resolution enhancement factor, U : Upsampling by zero insertion, $W_b^{(i)}$: Back-warping to the reference grid.

ing window size large may lead to artifacts in final result, as correlation is valid for a local neighborhood. Realization of color consistency constraint is exhaustive as it is implemented in a moving window fashion for each pixel on the grid. Consistency projection is implemented with the following operation

$$P_c \left[\mathbf{y}_S^{(i)} \right] = \begin{cases} \mathbf{y}_G^{(i)} + \overline{\left(\mathbf{y}_S^{(i)} - \mathbf{y}_G^{(i)} \right)} + \mathbf{T}_c, & \text{if } \left(\mathbf{y}_S^{(i)} - \mathbf{y}_G^{(i)} \right) - \overline{\left(\mathbf{y}_S^{(i)} - \mathbf{y}_G^{(i)} \right)} > \mathbf{T}_c \\ \mathbf{y}_S^{(i)}, & \text{if } \left| \left(\mathbf{y}_S^{(i)} - \mathbf{y}_G^{(i)} \right) - \overline{\left(\mathbf{y}_S^{(i)} - \mathbf{y}_G^{(i)} \right)} \right| \leq \mathbf{T}_c \\ \mathbf{y}_G^{(i)} + \overline{\left(\mathbf{y}_S^{(i)} - \mathbf{y}_G^{(i)} \right)} - \mathbf{T}_c, & \text{if } \left(\mathbf{y}_S^{(i)} - \mathbf{y}_G^{(i)} \right) - \overline{\left(\mathbf{y}_S^{(i)} - \mathbf{y}_G^{(i)} \right)} < -\mathbf{T}_c \end{cases}$$

where $S = R, B$. The overall procedure of multi-frame demosaicking is provided in Algorithm 1.

D. Achieving Subpixel Resolution

Using multiple images, it is possible to achieve subpixel resolution. This process is known as super-resolution reconstruction. We consider two approaches for CFA-sampled images: *demosaicking and super-resolution* (DSR) and *only super-resolution* (OSR).

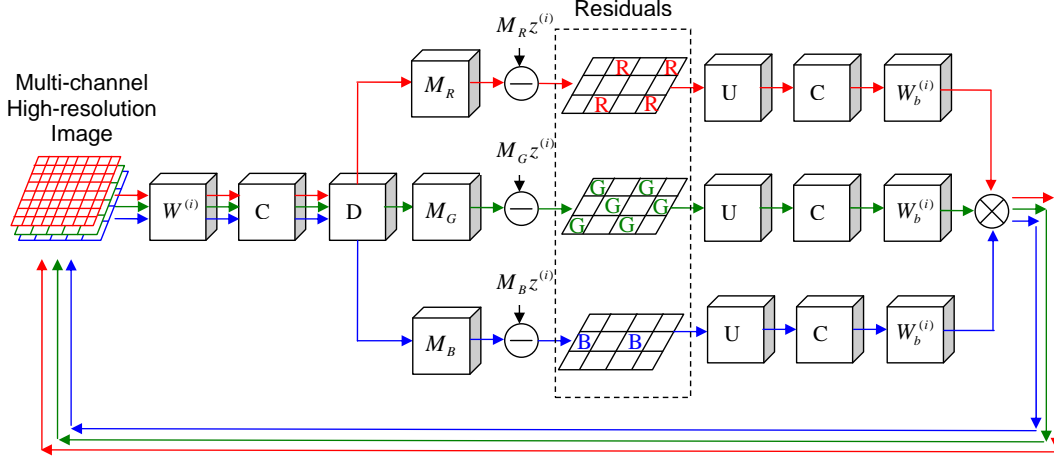


Fig. 34.: OSR reconstruction approach is illustrated. The algorithm starts with a high-resolution image estimate. Simulated observations are obtained by forward imaging operations, including the CFA sampling. The residuals are computed on Bayer pattern samples for each channel, and then back-projected. The notation in the figure is as follows. $W^{(i)}$: Spatial warping onto i th observation, C : Convolution with the PSF, D : Downsampling by the resolution enhancement factor, U : Upsampling by zero insertion, $W_b^{(i)}$: Back-warping to the reference grid.

1. Demosaicking Super-Resolution (DSR) Approach

In DSR approach, demosaicking is first performed on each frame, and then super-resolution reconstruction is applied to the demosaicked images. For demosaicking, either single-frame demosaicking or multi-frame demosaicking (as presented in this paper) can be applied. For super-resolution reconstruction, each color channel can be treated separately using any of the super-resolution algorithms in the literature. Figure 33 illustrates this idea for the iterated back-projection super-resolution algorithm [11]. Starting with an initial estimate for the high-resolution image, warping, blurring and downsampling operations are applied to produce a *simulated* observation. The residual between the *simulated* observation and the actual observation $\mathbf{y}_S^{(i)}$ is then back-projected. The back-projection operation includes upsampling by zero insertion, blurring, and back-warping. This is done for each channel separately. The process is repeated for a fixed number of iterations or until convergence.

2. Only Super-Resolution (OSR) Approach

The other approach is to perform demosaicking and super-resolution reconstruction jointly. That is, a high-resolution image is reconstructed from the input data $\mathbf{z}^{(i)}$ without estimating full-color low-resolution images $\mathbf{y}_S^{(i)}$. Figure 34 shows an illustration of this approach. The reconstruction process is based on the iterated back-projection algorithm: A high-resolution image estimate is warped, blurred, down-sampled, and CFA sampled to create a *simulated* observation. The residual between this *simulated* observation and the actual observation is computed, and then back-projected as explained previously. Note that in this approach the samples are more sparsely located than the DSR approach. This may create a problem (reconstruction artifacts) when there is not sufficient number of input images. The convolution kernels should have larger basis to compensate for the sparse sampling.

E. Experiments

1. Comparison of Single-Frame Demosaicking Algorithms

We would like to show the contribution of the spatio-intensity based constraint set to the alternating projections framework of [77]. Figure 35 shows the merit of spatio-intensity constraint set. Mean square error with and without consistency constraint set is 16.85 and 16.90, and mean square error of [1] is 16.34. Although there is a slight decrease in objective error, proposed constraint set makes algorithm superior in visual reconstruction as shown in Figure 35 (d) compared to earlier method, which is shown in Figure 35 (c). Proposed method is also visually comparable to [1], whose result is given in Figure 35 (b).

We compare several demosaicking algorithms. These include single-frame demosaicking algorithms such as bilinear interpolation, edge-directed interpolation [2], alternating projections [77] and adaptive homogeneity directed demosaicking [1] in comparison with our proposed demosaicking algorithm. In this section proposed algorithm is applied using single image framework as existing demosaicking algorithms

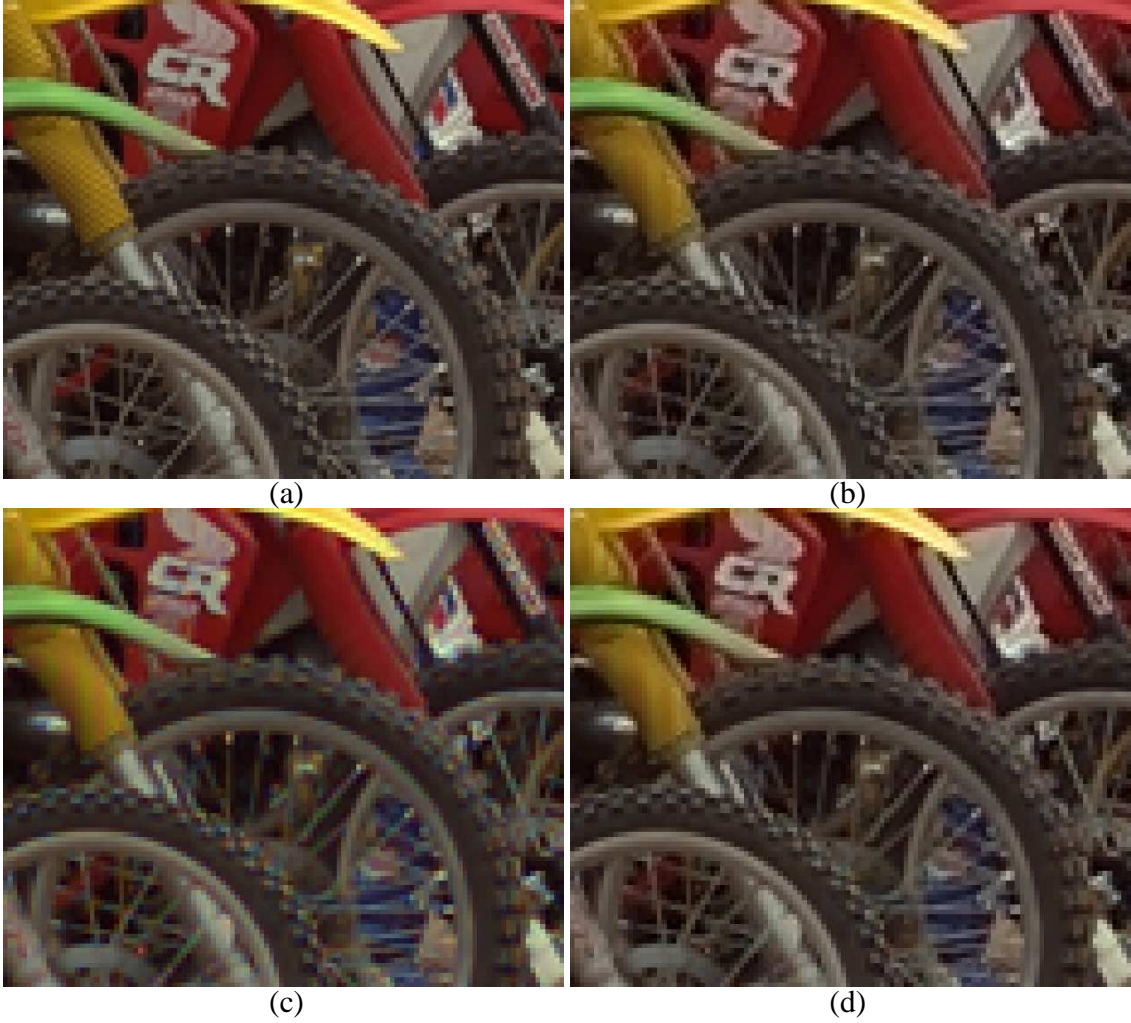


Fig. 35.: Merit of consistency constraint set. (a) Original image. (b) Adaptive homogeneity-directed demosaicking [1]. (c) Demosaicked without consistency constraint set in 4 iterations, $T_d = 5, T_o = 0$. (d) Demosaicked with consistency constraint set in 4 iterations, $\sigma_I = 1.41, \sigma_S = 1.41, T_d = 5, T_o = 0, T_c = 2$, domain window is $[5 \times 5]$.

are applied on single frame, and it is fair to make comparison on same basis. Multi-frame demosaicking in conjunction with the resolution enhancement will be performed in next section.

Another demosaicking comparison is provided in Figure 36. You can notice the merit of the consistency constraint set in Figure 36-(e). The result of the proposed method is visually similar to the method of [1]. Mean square error of [1], POCS with and without consistency constraint set are 14.80, 15.74, and 15.90, respectively for images in 36. The POCS framework is suitable to incorporate any additional constraint set. POCS code with all the constraint set applied is available in [68].

2. Super Resolution Experiments

In this section we compare several approaches for demosaicking and super-resolution reconstruction. We first captured a video sequence of length 21 using *CanonG5* digital camera. The size of each frame is 240×320 . These images are then downsampled by two and then further sampled with the Bayer pattern to produce the input data of size 120×160 . Figure 37 shows six of these input data. The motion parameters are estimated using feature-based image registration technique [84]. The motion parameters are estimated from the interpolated green channels due to its high sampling rate, and it turns out that images are related with affine transformation. Our second data set contains 50 images taken from a video sequence while camera is allowed to move only in x and y directions. These images are originally of size 120×160 and they are Bayer sampled to produce mosaic input patterns, 6 of which are shown in Figure 38.

We compared several super-resolution reconstruction approaches. We provide resolution enhancement results for two different data sets. In first experiment we used the Bayer pattern images given in Figure 37. Demosaicked and multi-frame enhanced results are given in Figure 39. Cropped regions of same results are given in Figure 40. In Figure 39 and Figure 40, (a), (b), (c), and (d) show single frame demosaicking

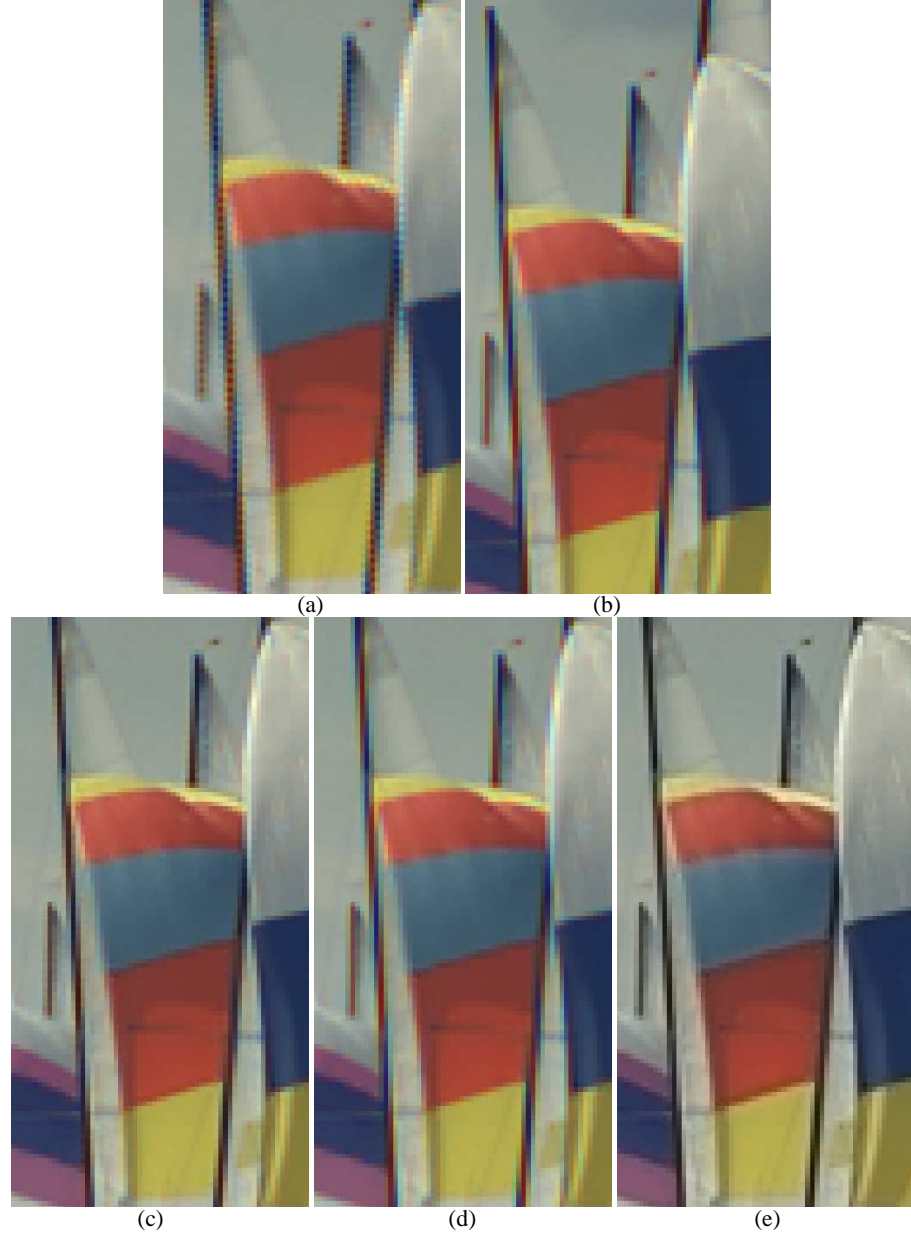


Fig. 36.: Single frame POCS comparisons. (a) Bilinear interpolation (b) Edge directed interpolation [2]. (c) Adaptive homogeneity-directed demosaicking [1]. (d) POCS based demosaicking without consistency set, $T_d = 5, T_o = 0$. (e) POCS based demosaicking with consistency set, $\sigma_I = 1.41, \sigma_S = 1.41, T_d = 5, T_o = 0, T_c = 2$, domain window is $[5 \times 5]$, demosaicking iteration number is 3.

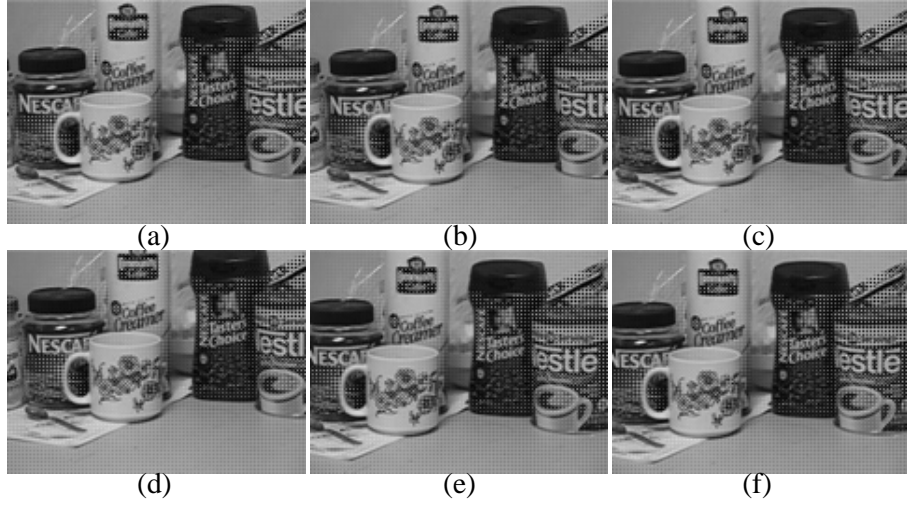


Fig. 37.: First data set. (a)-(f) Bayer pattern input images

results. (e) is multi-frame demosaicking result, which is visually superior and has less color artifacts compared to single frame demosaicking results. (f) is the OSR result using single frame demosaicking. (g) and (h) are DSR results without and with color consistency constraint sets using single frame demosaicking approach, respectively. (i) is resolution enhancement with consistency set using multi frame demosaicking approach. Note that border artifacts may occur in super resolution with multi-frame demosaicking due to the fact that backwarped residuals have non-overlapping regions. Final result should be evaluated in the intersecting region of warped residuals. These artifacts can be handled using only single frame demosaicking in non-overlapping parts. This approach is not implemented as it is out of the scope of the paper.

As seen in these figures, OSR has worse performance among multi-frame approaches. DSR with spatio-intensity outperforms others. Although multi-frame demosaicking approach outperforms single-frame methods, it does not have any significant contribution in case of super-resolution enhancement due to averaging effect. Comparing Figure 40 (g) and (h) and shows that spatio-intensity constraint removes the zipper artifacts. In these initial experiments, the reconstruction parameters were

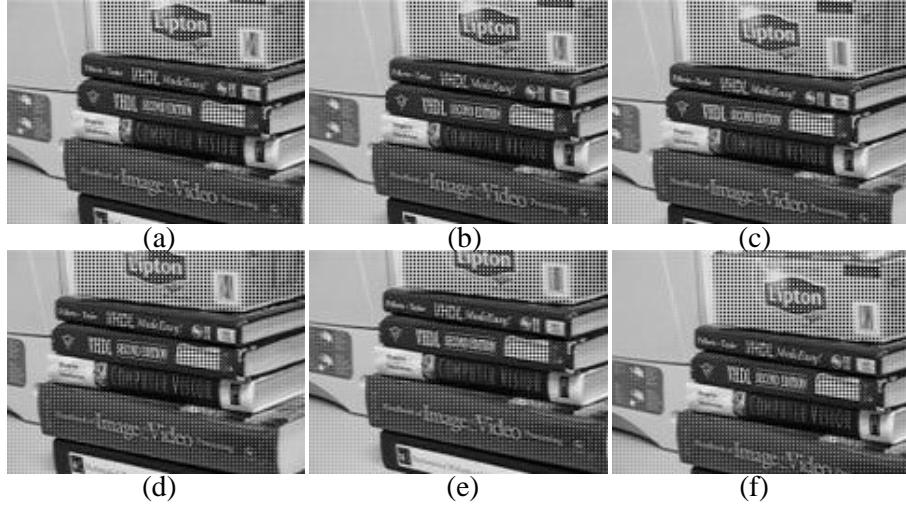


Fig. 38.: Second data set. (a)-(f) Bayer pattern input images

chosen by trial-and-error. As a future work, we will look into the parameter selection issue in more detail.

For second data set we provide comparison with the work of [3], which performs demosaicking and super-resolution jointly, using default parameters provided in their software. Results for this data set is given in Figure 41. Multi-frame approach outperforms the single-frame method as results indicate. Merit of spatio-intensity can be seen in Figure 41 (h) as zipper effects are less compared to Figure 41 (g). Although work of [3] produces results with less noise, it contains jagged artifacts as shown in Figure 41 (f). Matlab user interface of our DSR is provided in [68] to conduct DSR experiments using proposed constraint sets.

F. Conclusions

In this section, we present a POCS-based framework for resolution enhancement of Bayer-sampled video sequences. We defined three constraint sets to be used in single or multi-frame demosaicking. Additional constraint sets can be incorporated easily within the POCS framework. Multi-frame demosaicking algorithm performed better than single-frame demosaicking algorithms. We also investigated two different approaches for super-resolution reconstruction. The DSR approach performs demo-

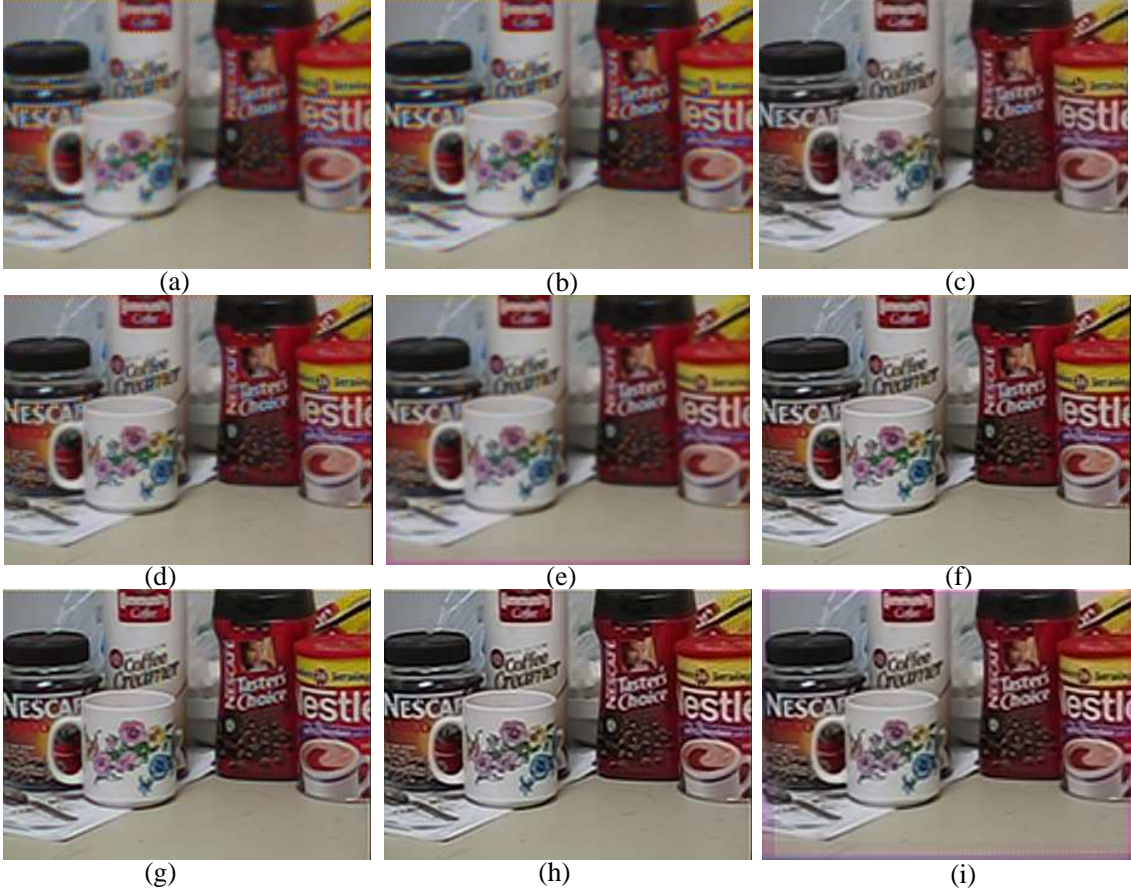


Fig. 39.: Demosaicking and super-resolution results. (a) Single frame demosaicking by bilinear interpolation. (b) Single frame demosaicking by edge directed interpolation [2]. (c) Single frame demosaicking by adaptive homogeneity directed interpolation [1]. (d) Proposed single frame demosaicking by all constraints applied, $\sigma_I = 1.41, \sigma_S = 1.41, T_d = 5, T_o = 0, T_c = 2$, domain window is $[5 \times 5]$, number of iterations is 3. (e) Multi-frame demosaicking, all constraints applied. Demosaicking is applied in 3 iterations, $\sigma_I = 1.41, \sigma_S = 1.41, T_d = 5, T_o = 0, T_c = 2$, domain window is $[5 \times 5]$. (f) OSR result using single frame demosaicking in 3 iterations. (g) DSR without consistency set using single frame demosaicking in 3 iterations, $T_d = 5, T_o = 0$. (h) DSR with all constraint sets applied in 3 iterations, $\sigma_I = 1.41, \sigma_S = 1.41, T_d = 5, T_o = 0, T_c = 2$, domain window is $[5 \times 5]$. (i) DSR using multi frame demosaicking with all constraint sets applied in 3 iterations, $\sigma_I = 1.41, \sigma_S = 1.41, T_d = 5, T_o = 0, T_c = 2$, domain window is $[5 \times 5]$. Iteration number for super resolution is set to 3 during the experiment.



Fig. 40.: Demosaicking and super-resolution results. (a) Single frame demosaicking by bilinear interpolation. (b) Single frame demosaicking by edge directed interpolation [2]. (c) Single frame demosaicking by adaptive homogeneity directed interpolation [1]. (d) Proposed single frame demosaicking by all constraints applied, $\sigma_I = 1.41, \sigma_S = 1.41, T_d = 5, T_o = 0, T_c = 2$, domain window is $[5 \times 5]$, iteration number is 3. (e) Multi-frame demosaicking, all constraints applied. Demosaicking is applied in 3 iterations, $\sigma_I = 1.41, \sigma_S = 1.41, T_d = 5, T_o = 0, T_c = 2$, domain window is $[5 \times 5]$. (f) OSR result using single frame demosaicking in 3 iterations. (g) DSR without consistency set using single frame demosaicking in 3 iterations, $T_d = 5, T_o = 0$. (h) DSR with all constraint sets applied in 3 iterations, $\sigma_I = 1.41, \sigma_S = 1.41, T_d = 5, T_o = 0, T_c = 2$, domain window is $[5 \times 5]$. (i) DSR using multi frame demosaicking with all constraint sets applied in 3 iterations, $\sigma_I = 1.41, \sigma_S = 1.41, T_d = 5, T_o = 0, T_c = 2$, domain window is $[5 \times 5]$. Iteration number for super resolution is kept 3 during the experiment.



Fig. 41.: Demosaicking and super-resolution results. (a) Single frame demosaicking by bilinear interpolation. (b) Single frame demosaicking by edge directed interpolation [2]. (c) Single frame demosaicking by adaptive homogeneity directed interpolation [1]. (d) Proposed single frame demosaicking by all constraints applied, $\sigma_I = 1.41, \sigma_S = 1.41, T_d = 5, T_o = 0, T_c = 2$, domain window is $[5 \times 5]$, iteration number is 3. (e) Multi-frame demosaicking, all constraints applied. Demosaicking is applied in 3 iterations, $\sigma_I = 1.41, \sigma_S = 1.41, T_d = 5, T_o = 0, T_c = 2$, domain window is $[5 \times 5]$. (f) OSR result using single frame demosaicking in 3 iterations. (g) Multi-frame demosaicking and super resolution of [3] with default parameters. (h) DSR using single frame demosaicking without consistency set in 3 iterations, $T_d = 5, T_o = 0$. (i) DSR using single frame demosaicking with all consistency sets applied in 3 iterations, $\sigma_I = 1.41, \sigma_S = 1.41, T_d = 5, T_o = 0, T_c = 2$, domain window is $[5 \times 5]$. (j) DSR using multi frame demosaicking with all constraint sets applied in 3 iterations, $\sigma_I = 1.41, \sigma_S = 1.41, T_d = 5, T_o = 0, T_c = 2$, domain window is $[5 \times 5]$

saicking and reconstruction of images together. The OSR approach performed worse than the DSR approach because no inter-channel correlation was utilized. In OSR approach, image is reconstructed in an iterated back-projection framework without demosaicking. The latter method is provided for comparison purposes. Another important observation is that when super resolution is applied, it did not matter much whether the demosaicking was performed using the single-frame or multi-frame approach.

Algorithm 1 Multi-Frame Demosaicking

Ensure: Total number of images are K

Require: Initialize parameters

- Set the maximum number of iterations (N)
- Set threshold values ($\mathbf{T}_d, \mathbf{T}_o, \mathbf{T}_c$)
- Set spatio-intensity neighborhood parameters ($\tau_S, \tau_I, \sigma_S, \sigma_I$)

1: **for** $i = 1 : K$ **do**

2: Apply single-frame demosaicking to each input data $\mathbf{z}^{(i)}$ to obtain $\mathbf{y}_S^{(i)}$

3: Estimate and save motion parameters using the interpolated green channels $\mathbf{y}_G^{(i)}$

4: **end for**

5: **for** $i = 1 : K$ **do**

6: Obtain the average image $\bar{\mathbf{y}}_S^{(i)}$ by warping all input images $\mathbf{y}_S^{(1)}, \mathbf{y}_S^{(2)}, \dots, \mathbf{y}_S^{(K)}$

7: $\mathbf{y}_G^{(i)} \leftarrow \bar{\mathbf{y}}_G^{(i)}$

8: **for** $n = 1 : N$ **do**

9: $\mathbf{y}_S^{(i)} \leftarrow P_d[\mathbf{y}_S^{(i)}]$, for $S = R, B$

10: $\mathbf{y}_S^{(i)} \leftarrow P_o[\mathbf{y}_S^{(i)}]$, for $S = R, B$

11: $\mathbf{y}_S^{(i)} \leftarrow P_c[\mathbf{y}_S^{(i)}]$, for $S = R, B$

12: **end for**

13: **end for**

CHAPTER V

CONCLUSIONS

Digital camera industry had the tendency to squeeze more number of pixels into sensor array for the last decade. This approach has been adapted not only to boost sharpness but also motivated by a competitive marketing point of view. Number of pixels in a camera, which is mistakenly conjectured as indication of sharpness, had been the leading decision factor for consumers. However, decrease in size of sensor site leads to increase in shot noise and a loss in dynamic range due to less photon collecting capability. This phenomenon led camera producers reconsider visual quality instead of struggling to boost up specifications. As a consequence, dynamic range criterion increased its rank among visual quality assessment methods.

Visual quality captured by a camera is degraded through imaging pipeline, during network transmission or due to compression scheme being applied. Image enhancement techniques aims to compensate for aforementioned degradations. Although joint spatial resolution and dynamic range increase has become a concern for camera producers, this technique has not been applied by image enhancement community. Multi-frame image fusion has been either used for spatial resolution enhancement or to increase dynamic range. Main contribution of this thesis was to perform joint tonal range improvement and spatial resolution enhancement over degraded image sequences. This challenging goal forced us to revise and redesign fundamental processes in image processing that are taken for granted. We believe that algorithms devised during our research will be vital for image enhancement and computer vision tasks.

Image formation pipeline is modelled to relate ground truth value to degraded images in Chapter 1. Photometric modelling has not been incorporated into image acquisition model up to now. Precision of our image model improves the success

of photometric and geometric image registration, which are fundamental for super-resolution dynamic range enhancement.

Chapter 2 presents contrast invariant feature transform (CIFT) to increase repeatability rate of salient points among different images under illumination change. CIFT is applied as feature detection algorithm for registering a high dynamic range image sequence. Repeatability rates of regular feature detectors fall drastically when video sequence consists of under-exposed and over-exposed images. Proposed feature detector is contrast invariant compared to the existing work in literature. CIFT considers possible contrast variations in addition to the spatial variations. Repeatability rate of CIFT is compared to state of art corner detection algorithms on real and synthetic images. Although CIFT is used for corner detection in this thesis, it can be extended for detecting other features i.e edge, contours [85].

Chapter 3 establishes the framework for super resolution under photometric diversity. The only photometric alignment in super resolution (SR) literature has been a gain and offset model to our knowledge [55]. Our proposed photometric alignment is based on camera response function and yields superior results compared to [55]. Bayesian (MAP) framework is used for reconstructing an SR image with high dynamic range (HDR) in irradiance domain [86]. We also demonstrate that it is possible to preserve the photometric range of reference image by computing intensity mapping function from camera response function and selecting a proper weighting function. SR-HDR and saturation elimination using weighting function is demonstrated on real data sets. Projection onto convex sets (POCS) was also successfully applied to SR-HDR image reconstruction in our various works [87] [56].

Chapter 4 focused on super resolution using color filter array (CFA) sampled images. CFA interpolation imposes challenges on image enhancement by introducing color artifacts. Correlation among color channels are incorporated into SR framework to produce high resolution images without demosaicking artifacts. POCS based

demosaiicking of Gunturk et. al is extended to multi-frame by incorporating a color constancy constraint set [69]. Color constancy constraint forces pixels with *spatio-intensity neighborhood* to be used in SR reconstruction with bilateral filtering approach. Color constancy together with detail and observation constraint sets produces sharper results with less color artifacts [88] [89].

Although wide variety of problems in SR were addressed in this thesis, there is still much to be investigated in area of multi-frame image enhancement. Geometric registration methods in SR are limited to parametric models, which limit the application areas. A geometric alignment procedure to handle occlusion and generic movements have not been proposed yet. Computational cost of SR algorithms should also be reduced to make existing algorithms suitable for real-time applications. Parallel programming will contribute heavily to the efficiency of SR algorithms in future. Prior selection is also an area to be investigated for regularization in Bayesian framework. Priors that consider the compression scheme and network conditions will lead to results with superior visual quality. Blur identification is another topic that is crucial for successful reconstruction. Estimated blur parameters can be think of as prior information to be used by reconstruction.

REFERENCES

- [1] K. Hirakawa and T.W. Parks, "Adaptive homogeneity-directed demosaicing algorithm," *IEEE Trans. Image Processing*, vol. 14, no. 3, pp. 360–369, March 2005.
- [2] C. A. Laroche and M. A. Prescott, "Apparatus and method for adaptively interpolating a full color image utilizing chrominance gradients," *U.S. Patent 5,373,322*, 1994.
- [3] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *IEEE Trans. Image Processing*, vol. 15, no. 1, pp. 141–159, January 2006.
- [4] T. Mitsunaga and S.K. Nayar, "Radiometric self calibration," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, June 1999, vol. 1, pp. 374–380.
- [5] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, May 2003.
- [6] R. Y. Tsai and T. S. Huang, *Multiframe Image Restoration and Registration, In: Advances in Computer Vision and Image Processing*, ed. T.S.Huang. Greenwich, CT, JAI Press, 1984.
- [7] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframe," *IEEE Trans. Acoust. Speech Sign. Proc.*, vol. 38, pp. 1013–1027, June 1990.
- [8] S. P. Kim and W. Su, "Recursive high-resolution reconstruction of blurred multiframe images," *IEEE Trans. Image Processing*, vol. 2, pp. 534–539, October 1993.
- [9] N. K. Bose, H. C. Kim, and H. M. Valenzuela, "Recursive total least squares algorithm for image reconstruction from noisy, undersampled frames," *Multidimensional Systems and Signal Processing*, vol. 4, pp. 253–268, July 1993.
- [10] S.H. Rhee and M.G. Kang, "Discrete cosine transform based regularized high-resolution image reconstruction algorithm," *Optical Engineering*, vol. 38, no. 8, pp. 1348–1356, April 1999.
- [11] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, pp. 231–239, May 1991.
- [12] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion, and transparency," *J. of Visual Communications and Image Representation*, vol. 4, pp. 324–335, December 1993.

- [13] B. C. Tom and A. K. Katsaggelos, "Reconstruction of a high resolution image by simultaneous registration, restoration, and interpolation of low-resolution images," in *Proc. IEEE Int. Conf. Image Processing*, October 1995, vol. 2, pp. 539–542.
- [14] H. Engl, M. Hanke, and A. Neubauer., *Regularization of Inverse Problems*, Dordrecht: Kluwer Academic Publishers, 1996.
- [15] C. Groetsch., *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman, 1984.
- [16] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1621–1633, December 1997.
- [17] P. Cheeseman, B. Kanefsky, R. Hanson, and J. Stutz, "Super-resolved surface reconstruction from multiple images," Tech. Rep. FIA-94-12, NASA Ames Research Center, December 1994.
- [18] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing*, vol. 5, no. 6, pp. 996–1011, June 1996.
- [19] D. Capel and A. Zisserman, "Super-resolution from multiple views using learnt image models," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, December 2001, vol. 2, pp. 627–634.
- [20] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes, and R. M. Mersereau, "Eigenface-domain super-resolution reconstruction," *IEEE Trans. Image Processing*, vol. 12, no. 5, pp. 597–606, May 2003.
- [21] N.K.Bose, S. Lertrattanapanich, and J. Koo, "Advances in superresolution using l-curve," in *The 2001 IEEE International Symposium on Circuit and Systems.*, May 2001, vol. 2, pp. 433–436.
- [22] N.Nguyen, G. Golub, and P. Milanfar, "Blind restoration/superresolution with generalized cross-validation using gauss-type quadrature rules," in *Conference Record of the Third-Third Asilomar Conference on Signals, Systems, and Computers*, October 1999, vol. 2, pp. 1257–1261.
- [23] A. Zomet and S. Peleg, *Super-Resolution from Multiple Images having Arbitrary Mutual Information*, Boston: Kluwer Academic Publishers, 2001.
- [24] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," *IEEE Trans. Image Processing*, October 2004.
- [25] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *J. of the Optical Society of America*, vol. 6, no. 11, pp. 1715–1726, 1989.

- [26] 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. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, March 1992, pp. 169–172.
- [27] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Processing*, vol. 6, no. 8, pp. 1064–1076, August 1997.
- [28] P. E. Eren, M. I. Sezan, and A. M. Tekalp, "Robust, object-based high-resolution image reconstruction from low-resolution," *IEEE Trans. Image Processing*, vol. 6, no. 10, pp. 1446–1451, October 1997.
- [29] A. J. Patti and Y. Altunbasak, "Artifact reduction for set theoretic super resolution image reconstruction with edge adaptive constraints and higher-order interpolants," *IEEE Trans. Image Processing*, vol. 10, no. 1, pp. 179–186, January 2001.
- [30] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy and undersampled measured images," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1646–1658, December 1997.
- [31] Stanley J. Reeves and Russell M. Mersereau, "Blur identification by the method of generalized cross-validation," *IEEE Trans. Image Processing*, vol. 1, pp. 301–311, 1992.
- [32] A. W. M. van Eekeren, K. Schutte, O. R. Oudegeest, and L. J. van Vliet, "Performance evaluation of super-resolution reconstruction methods on real-world data," *EURASIP J. Appl. Signal Process.*, April 2007.
- [33] S. Battiato, A. Castorina, and M. Mancuso, "High dynamic range imaging for digital still camera: an overview," *Journal of Electronic Imaging*, vol. 12, no. 3, pp. 459–468, July 2003.
- [34] P. E. Debevec and J. Malik, "Modeling and rendering architecture from photographs," in *Proc. of the ACM SIGGRAPH*, 1997, pp. 369–378.
- [35] Y. Tsin, V. Ramesh, and T. Kanade, "Statistical calibration of the ccd imaging process," *Proc. Int. Conf. Computer Vision*, vol. 1, pp. 480–487, 2001.
- [36] S. Mann, C. Manders, and J. Fung, "The lightspace change constraint equation (lcce) with practical application to estimation of the projectivity+gain transformation between multiple pictures of the same subject matter," *Image and Multidimensional Signal Processing: Image and Video Analysis*, vol. 1, April 2003.
- [37] F.M. Candocia, "Simultaneous homographic and comparametric alignment of multiple exposure-adjusted pictures of the same scene," *IEEE Trans. Image Processing*, vol. 12, no. 12, pp. 1485–1494, 2003.

- [38] S. Mann, “Comparametric equations with practical applications in quantigraphic image processing,” *IEEE Trans. Image Processing*, vol. 9, no. 8, pp. 1389–1406, August 2000.
- [39] J. L. Barron, D. J. Fleet, S. S. Beauchemin, and T. A. Burkitt, “Performance of optical flow techniques,” *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 92, pp. 236–242.
- [40] P. Thévenaz, U.E. Ruttimann, and M. Unser, “A pyramid approach to subpixel registration based on intensity,” *IEEE Trans. Image Processing*, vol. 7, no. 1, pp. 27–41, January 1998.
- [41] D. Capel and A. Zisserman, “Computer vision applied to super resolution,” *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 75–86, May 2003.
- [42] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [43] C. J. Harris and M. Stephens, “A combined corner and edge detector,” in *Proceedings of the 4th Alvey Vision Conference*, 1988, pp. 147–151.
- [44] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Comm. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [45] H. Moravec, “Obstacle avoidance and navigation in the real world by a seeing robot rover,” doctoral dissertation Technical Report CMU-RI-TR-80-03, Robotics Institute, Carnegie Mellon University, September 1980.
- [46] S.M. Smith and J.M. Brady, “Susan: A new approach to low-level image-processing,” *International Journal of Computer Vision*, vol. 23, no. 1, pp. 45–78, May 1997.
- [47] P. Kovesi, “Image features from phase congruency,” *Videre*, vol. 1, no. 3, pp. 1–26, 1999.
- [48] M.C. Morrone, J.R. Ross, D.C. Burr, and R.A. Owens, “Mach bands are phase dependent,” *Nature*, vol. 324, no. 3, pp. 250–253, 1986.
- [49] K. Mikolajczyk and C. Schmid, “Scale and affine invariant interest point detectors,” *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.
- [50] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [51] D.G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, November 2004.

- [52] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, ,” in *High dynamic range imaging: acquisition, display and image-based lighting*. 2006, Morgan Kaufmann.
- [53] S. Mann and R. W. Picard, “Being undigital with digital cameras: extending dynamic range by combining differently exposed pictures,” in *IS&T 48th Annual Conference*, May 1995.
- [54] M. A. Robertson, S. Borman, and R. L. Stevenson, “Dynamic range improvement through multiple exposures,” in *Proc. IEEE Int. Conf. Image Processing*, 1999, vol. 3, pp. 159–163.
- [55] D. Capel, *Image mosaicing and super resolution*, Springer, 2004.
- [56] M. Gevrekci and B. K. Gunturk, “Superresolution under photometric diversity of images,” *IEEE Signal Processing Letters*, vol. 13, pp. 197–200, April 2006.
- [57] Y.Y Litvinov, A. Schechner, “Addressing radiometric nonidealities: A unified framework,” *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 2, no. 12, pp. 52–59, 2005.
- [58] D. Hasler and S. Susstrunk, “Mapping colour in image stitching applications,” vol. 15, pp. 65–90, 2004.
- [59] K. Ikeuchi S. K. Nayar and T. Kanade, “Shape from interreflections,” *Int’ J. Computer Vision*, vol. 6, no. 3, pp. 173–195, August 1991.
- [60] M. D. Grossberg and S. K. Nayar, “Determining the camera response from images: what is knowable?,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1455–1467, November 2003.
- [61] S. Mann, C. Manders, and J. Fung, “Painting with looks: photographic images from video using quantimetric processing,” pp. 117–126, 2002.
- [62] S. Mann and R. Mann, “Quantigraphic imaging: estimating the camera response and exposures from differently exposed images,” vol. 1, pp. 842–849, 2001.
- [63] T. Mitsunaga and S.K. Nayar, “Radiometric self calibration,” in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, June 1999, vol. 1, pp. 374–380.
- [64] R. Fattal, D. Lischinski, and M. Werman, “Gradient domain high dynamic range compression,” *Proc. ACM SIGGRAPH*, July 2002.
- [65] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, “High dynamic range video,” *ACM Transactions on Graphics*, vol. 23, no. 3, 2003.
- [66] G. Ward, “Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures,” *Journal of Graphics Tools*, vol. 8, no. 2, pp. 17–30, 2003.

- [67] F. M. Candocia, "Joint registering images in domain and range by piecewise linear comparametric analysis," *IEEE Trans. Image Processing*, vol. 12, no. 4, pp. 409–419, 2003.
- [68] M. Gevrekci and B. K. Gunturk, "Matlab user interface for super resolution image reconstruction for illumination varying and bayern images, dept. of electrical and computer engineering, louisiana state university, september 2008. <http://www.ece.lsu.edu/ipl>," 2008.
- [69] B. K. Gunturk, J. Glothzbach, Y. Altunbasak, R. W. Schafer, and R. M. Mersereau, "Demosaiicing: Color filter array interpolation," *IEEE Signal Processing Magazine*, vol. 22, no. 1, pp. 44–54, January 2005.
- [70] R. H. Hibbard, "Apparatus and method for adaptively interpolating a full color image utilizing luminance gradients," *U.S. Patent 5,382,976*, 1995.
- [71] J. E. Adams, "Interactions between color plane interpolation and other image processing functions in electronic photography," in *Proc. SPIE*, 1995, vol. 2416, pp. 144–151.
- [72] W. T. Freeman, "Method and apparatus for reconstructing missing color samples," *U.S. Patent 4,774,565*, 1988.
- [73] R. Kimmel, "Demosaiicing: image reconstruction from ccd samples," *IEEE Trans. Image Processing*, vol. 8, pp. 1221–1228, 1999.
- [74] S.-C. Pei and I.-K. Tam, "Effective color interpolation in ccd color filter arrays using signal correlation," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 6, pp. 503–513, June 2003.
- [75] J. E. Adams and J. F. Hamilton, "Design of practical color filter array interpolation algorithms for digital cameras," in *Proc. SPIE*, 1997, vol. 3028, pp. 117–125.
- [76] J. W. Glotzbach, R. W. Schafer, and K. Illgner, "A method of color filter array interpolation with alias cancellation properties," in *Proc. IEEE Int. Conf. Image Processing*, 2001, vol. 1, pp. 141–144.
- [77] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Multiframe resolution-enhancement methods for compressed video," *Signal Processing Letters*, vol. 9, no. 6, pp. 170–174, June 2002.
- [78] J. Mukherjee, R. Parthasarathi, and S. Goyal, "Markov random field processing for color demosaicing," *Pattern Recognition Letters*, vol. 22, no. 3-4, pp. 339–351, March 2001.
- [79] O. Kapah and H. Z. Hel-Or, "Demosaiicing using artificial neural networks," in *Proc. SPIE*, 2000, vol. 3962, pp. 112–120.

- [80] K. Hirakawa and T. W. Parks, "Adaptive homogeneity-directed demosaicing algorithm," *Proc. IEEE Int. Conf. Image Processing*, vol. 3, pp. 669–672, 2003.
- [81] T. Gotoh and M. Okutomi, "Direct super-resolution and registration using raw cfa images," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 600–607, June–July 2004.
- [82] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. of the IEEE*, vol. 81, no. 2, pp. 182–208, February 1993.
- [83] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the 1998 IEEE International Conference on Computer Vision*, January 1998, pp. 839–846.
- [84] D. Capel and A. Zisserman, "Super-resolution enhancement of text image sequences," in *Proc. IEEE Int. Conf. Pattern Recognition*, 2000, vol. 1, pp. 600–605.
- [85] M. Gevrekci and B. K. Gunturk, "On geometric and photometric registration of images," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, April 2007, vol. 1, pp. 1261–1264.
- [86] Murat Gevrekci and Bahadır K. Gunturk, "Superresolution under photometric diversity of images," *EURASIP J. Appl. Signal Process.*, vol. 1, pp. 205–267, January 2007.
- [87] M. Gevrekci and B. K. Gunturk, "Image acquisition modeling for super-resolution reconstruction," in *Proc. IEEE Int. Conf. Image Processing*, 9 2005, vol. 2, pp. 1058–1061.
- [88] Murat Gevrekci, Bahadır K. Gunturk, and Yucel Altunbasak, "Restoration of bayer-sampled image sequences," *Oxford Journals*, January 2007.
- [89] M. Gevrekci, B. K. Gunturk, and Y. Altunbasak, "Pocs-based restoration of bayer-sampled image sequences," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, April 2007, vol. 1, pp. 753–756.

APPENDIX A

ITERATIVE IMAGE ENHANCEMENT

Steepest descent (SD) and conjugate gradient (CG) solution of SR image are provided in this chapter. MAP solution of SR can be obtained by following minimization.

$$\mathbf{x}_{map} = \arg \min_f \gamma^2 \|\mathbf{L}\mathbf{x}\|_2^2 + \sum_{k=1}^N \|\mathbf{y}_k - \mathbf{H}_k\mathbf{x}\|_2^2, \quad (\text{A.1})$$

where L is the discrete approximation of a derivative operator, Laplacian or identity matrix. Here regularization term provides a smoother solution for HR. Regularization parameter (γ) should be chosen carefully as smoothing operations acting on edges remove the sharpness. Taking the derivative of Eq .A.1 with respect to \mathbf{x}

$$\left[\sum_{k=1}^N \mathbf{H}_k^T \mathbf{H}_k + \gamma^2 \mathbf{L}^T \mathbf{L} \right] \mathbf{x} = \sum_{k=1}^N \mathbf{H}_k^T \mathbf{y}_k \quad (\text{A.2})$$

This equation can easily be solved using steepest descent (SD) algorithm. The $n + 1^{\text{th}}$ estimate of HR is as follows using SD:

$$\mathbf{x}^{n+1} = \mathbf{x}^n + \lambda_n \mathbf{p}_n, \quad (\text{A.3})$$

where $\mathbf{p}_n = \left[\sum_{k=1}^N \mathbf{H}_k^T (\mathbf{y}_k - \mathbf{H}_k \mathbf{x}^n) - \gamma^2 \mathbf{L}^T \mathbf{L} \mathbf{x}^n \right]$ and λ_n is the step size which determines the convergence. For quadratic forms step size is suggested as following by Capel[55]:

$$\lambda_n = \frac{\mathbf{p}_n^T \mathbf{p}_n}{\mathbf{p}_n^T \mathbf{C} \mathbf{p}_n}, \quad (\text{A.4})$$

where \mathbf{C} is the Hessian matrix of the quadratic form i.e $J(\mathbf{x}) = a + \mathbf{b}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{C} \mathbf{x}$. Step size in Eq .A.4 ensures that $J(\mathbf{x}^n + \lambda_n \mathbf{p}_n) < J(\mathbf{x}^n)$. SD is a slow iterative method in terms of convergence. Conjugate gradient (CG) is a fast method that provides convergence in a fast manner compared to SD. In CG, descent directions are chosen to be conjugate to the previous search directions, i.e $\mathbf{p}_n^T \mathbf{C} \mathbf{p}_j = 0$, for $\forall j$.

Convergence is guaranteed for a finite number of steps. Denoting $\nabla J(\mathbf{x}^n)$ as ∇J_n , new search direction becomes

$$\mathbf{p}_n = \nabla J_n + \left(\frac{\nabla J_n^T \nabla J_n}{\nabla J_{n-1}^T \nabla J_{n-1}} \right) \mathbf{p}_{n-1} \quad (\text{A.5})$$

Zomet and Peleg gives a detailed description of CG solution for SR image [23]. The outline of the CG algorithm for solving $\mathbf{y} = \mathbf{H}\mathbf{x}^0$ is given below for the sake of completeness.

1. Initialize maximum iteration number ($mMax$), and closeness threshold (ϵ). Form the initial estimate for SR (\mathbf{x}^0).
2. $\mathbf{r} = \mathbf{y} - \mathbf{H}\mathbf{x}^0$, $\rho_0 = \|\mathbf{r}\|^2$, $m = 1$
3. Do While $\sqrt{\rho_{m-1}} > \epsilon\|\mathbf{y}\|^2$ and $m < mMax$
 - (a) if $m = 1$ then $\mathbf{p} = \mathbf{r}$
else $\beta = \frac{\rho_{m-1}}{\rho_{m-2}}$ and $\mathbf{p} = \mathbf{r} + \beta\mathbf{p}$
 - (b) $\mathbf{w} = \mathbf{H}^T \mathbf{H}\mathbf{p}$
 - (c) $\alpha = \frac{\rho_{m-1}}{\mathbf{p}^T \mathbf{w}}$
 - (d) $\hat{\mathbf{f}}^m = \hat{\mathbf{f}}^m + \alpha\mathbf{p}$, $\mathbf{r} = \mathbf{r} - \alpha\mathbf{w}$, $\rho_m = \|\mathbf{r}\|^2$, $m = m + 1$.

Initial residual is found as $\mathbf{r} = \sum_{n=1}^N \mathbf{H}_n^T (\mathbf{y}_n - \mathbf{H}_n \mathbf{x}^0)$ and $\mathbf{w} = \sum_{n=1}^N \mathbf{H}_n^T \mathbf{H}_n \mathbf{p}$. Desired prior information can be incorporated in (d) step of the algorithm. The way we update SR image using residuals effect the convergence of both SD and CG. Residuals can be updated in either *incremental* or *batch* mode. In *incremental* mode SR image is updated by presenting each observation and its corresponding residual. This approach makes final SR result dependent on the order which observations are presented. In *batch* mode residual of each observation is computed using current estimate of SR image, and these residuals are accumulated. SR image is updated after all observations are presented. *Batch* mode provides superior results compared to *incremental* mode.

APPENDIX B

SUBBAND IMAGE DECOMPOSITION

Here, we define the subband decomposition notation. We use a set of filters to decompose an image into its frequency subbands and then reconstruct it back. The filter bank performs an undecimated wavelet transform, with $H_0(z)$ and $H_1(z)$ denoting low-pass and high-pass filters, respectively. These analysis filters ($H_0(z)$ and $H_1(z)$) constitute a perfect reconstruction filter bank with the synthesis filters $G_0(z)$ and $G_1(z)$. The perfect reconstruction condition can be written as:

$$H_0(z)G_0(z) + H_1(z)G_1(z) = 1. \quad (\text{B.1})$$

By denoting $h_0(\cdot)$ and $h_1(\cdot)$ as the impulse responses of $H_0(z)$ and $H_1(z)$, respectively, we can write the four subbands of a two-dimensional signal $S(n_1, n_2)$ as follows:

$$(\mathcal{W}_{LL}S)(n_1, n_2) = h_0(n_1) * [h_0(n_2) * S(n_1, n_2)] \quad (\text{B.2})$$

$$(\mathcal{W}_{HL}S)(n_1, n_2) = h_1(n_1) * [h_0(n_2) * S(n_1, n_2)] \quad (\text{B.3})$$

$$(\mathcal{W}_{LH}S)(n_1, n_2) = h_0(n_1) * [h_1(n_2) * S(n_1, n_2)] \quad (\text{B.4})$$

$$(\mathcal{W}_{HH}S)(n_1, n_2) = h_1(n_1) * [h_1(n_2) * S(n_1, n_2)], \quad (\text{B.5})$$

where $(\mathcal{W}_{LL}S)$ is the approximation subband, and $(\mathcal{W}_{HL}S)$, $(\mathcal{W}_{LH}S)$, $(\mathcal{W}_{HH}S)$ are the horizontal, vertical, and diagonal detail subbands, respectively.

We can similarly define the filtering operations in the synthesis stage of the filter bank. Letting $g_0(\cdot)$ and $g_1(\cdot)$ denote the impulse responses corresponding to $G_0(z)$ and $G_1(z)$, we can write the four filtering operations on a two-dimensional signal $X(n_1, n_2)$ as follows:

$$(\mathcal{U}_{LL}X)(n_1, n_2) = g_0(n_1) * [g_0(n_2) * X(n_1, n_2)] \quad (\text{B.6})$$

$$(\mathcal{U}_{HL}X)(n_1, n_2) = g_1(n_1) * [g_0(n_2) * X(n_1, n_2)] \quad (\text{B.7})$$

$$(\mathcal{U}_{LH}X)(n_1, n_2) = g_0(n_1) * [g_1(n_2) * X(n_1, n_2)] \quad (\text{B.8})$$

$$(\mathcal{U}_{HH}X)(n_1, n_2) = g_1(n_1) * [g_1(n_2) * X(n_1, n_2)], \quad (\text{B.9})$$

where \mathcal{U}_{LL} , \mathcal{U}_{HL} , \mathcal{U}_{LH} , \mathcal{U}_{HH} are the synthesis filtering operators. As stated earlier, these form a perfect reconstruction filter bank with the analysis filtering operators \mathcal{W}_{LL} , \mathcal{W}_{HL} , \mathcal{W}_{LH} , \mathcal{W}_{HH} :

$$\begin{aligned} S(n_1, n_2) = & \mathcal{U}_{LL}(\mathcal{W}_{LL}S)(n_1, n_2) + \mathcal{U}_{HL}(\mathcal{W}_{HL}S)(n_1, n_2) + \\ & + \mathcal{U}_{LH}(\mathcal{W}_{LH}S)(n_1, n_2) + \mathcal{U}_{HH}(\mathcal{W}_{HH}S)(n_1, n_2) \end{aligned} \quad (\text{B.10})$$

VITA

Lutfi Murat Gevrekci was born in Ankara, Turkiye, in April, 1982, the son of Huseyin Gevrekci and Munire Gevrekci. After completing Ataturk Anatolian High School, Ankara, Turkiye, in 2000, he entered Bilkent University in Ankara, Turkiye, where he received his Bachelor of Science Degree in May, 2004. He joined Louisiana State University graduate program in Baton Rouge, Louisiana, in July, 2004, and received his Master of Science Degree in July, 2007. During his graduate studies, he was an intern for Microsoft and Label Vision Systems as an image processing engineer. In January, 2008, he started working as a computer vision engineer for Label Vision Systems in Atlanta.