

PROJECTIVE IMAGE RESTORATION USING SPARSITY REGULARIZATION

N. Anantrasirichai, J. Burn and David R. Bull

Bristol Vision Institute, University of Bristol, Bristol BS8 1UB, UK

ABSTRACT

This paper presents a method of image restoration for projective ground images which lie on a projection orthogonal to the camera axis. The ground images are initially transformed using homography, and then the proposed image restoration is applied. The process is performed in the dual-tree complex wavelet transform domain in conjunction with L0 reweighting and L2 minimisation (L_0RL_2) employed to solve this ill-posed problem. We also propose instant estimation of a blur kernel arising from the projective transform and the subsequent interpolation of sparse data. Subjective results show significant improvement of image quality. Furthermore, classification of surface type at various distances (evaluated using a support vector machine classifier) is also improved for the images restored using our proposed algorithm.

Index Terms— image restoration, projective transform, DT-CWT

1. INTRODUCTION

Numerous scenarios exist where it is necessary or advantageous to classify surface material at a distance from a moving forward-facing camera. Examples include the use of image based sensors for assessing and predicting terrain type in association with the control or navigation of autonomous vehicles. In such cases, the ground ahead of the camera appears as a projective plane on the image - an ordinary plane with additional ‘points at infinity’ where parallel lines intersect. The ground plane in the three-dimensional world coordinate system is projected onto the image plane by means of straight visual rays from the point in space to the optical centre. This process can be described mathematically by a plane-to-plane homography \mathbf{H} (Projective Transformation) satisfying the relationship $\mathbf{x} = \mathbf{H}\mathbf{X}$, where \mathbf{x} and \mathbf{X} are points on the image plane and the world plane, respectively.

The homography of such a plane can be computed simply by knowing the relative position of four points on the scene plane and their corresponding positions in the image. An example showing a ground image and its reconstruction is shown in Fig. 1. The rectification of the surface reveals instances and extensions of the pattern which are very difficult to discern by eye in the original image. Homography is an irreversible transformation as transforming the projective

image back to the rectified view using traditional bilinear or bicubic interpolation produces noticeable artifacts, particularly at far distances. This ill-posed problem can be expressed with a matrix-vector multiplication as $\mathbf{I}_{obv} = \mathbf{D}\mathbf{I}_{idl} + \varepsilon$, where \mathbf{I}_{obv} and \mathbf{I}_{idl} are vectors containing the observed and ideal images, respectively. Matrix \mathbf{D} represents projective transformation, anti-aliasing and blur, while ε represents noise. Various approaches have attempted to solve this problem, generally by modelling it as a point spread function (PSF), where \mathbf{D} is considered as a convolution matrix, and then employing deconvolution with an iterative process to estimate \mathbf{I}_{idl} [1, 2].

The ground image obtained from the camera can be considered as an under-sampled observation of its rectified view in Fig. 1. As we move further away from the camera, the effective sampling rate decreases, causing increasing difficulty in approximating or assessing the surface texture. In this paper, we restore the projective image using a model-based recovery method which comprises convolution and subsampling. This problem is solved using L0 reweighted-L2 minimisation (L_0RL_2) in the wavelet domain [3]. The wavelet transform is employed because its decompositions provide image information about structure, strong persistence across scales and compressible properties. These properties are very important for establishing sparsity regularization [4].

The main challenge associated with this restoration is the estimation of the PSF which is generally unknown. In this paper, we address PSF estimation by exploiting information already present in the image. For example, in the rectified ground image, the area close to the camera is invariably sharper than other image regions further away (assuming of course that an appropriate aperture is used when acquiring the image). We compare this near area with the current region under analysis at a particular distance, in order to estimate the current blur kernel. The size of the kernel is equal to the area on the rectified image corresponding to the projection of the one step pixel distance in the observed image.

The remaining part of this paper is organised as follows. The proposed scheme for projective image restoration is described in detail in Section 2. The performance of the method is evaluated on a set of images in Section 3, both in terms of subjective performance and also in terms of classifier results enhanced using the restored data. Finally, Section 4 presents the conclusions of the study.

2. PROPOSED SCHEME

The proposed algorithm is performed in the wavelet domain which employs subband-dependent minimization and the dual-tree complex wavelet transform (DT-CWT) in an iterative Bayesian framework. The observation model for projective image restoration can be written as Eq. 1.

$$\mathbf{y} = \mathbf{A}\mathbf{H}\mathbf{B}\mathbf{x} + \varepsilon \quad (1)$$

where \mathbf{y} and \mathbf{x} are the observed ground image and the original projective (rectified) image, respectively. \mathbf{B} is a blur matrix, while \mathbf{H} and \mathbf{A} are a projective transform (homography) and an anti-aliasing matrix, respectively. We assume that ε is independent white noise so that we can get a maximum a posteriori probability (MAP) estimate of \mathbf{x} by minimising a cost function $C(\mathbf{x})$ as Eq. 2, where $\log(p(\mathbf{x}))$ describes prior information of image structure and λ_r is the regularization parameter.

$$C(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{H}\mathbf{B}\mathbf{x}\|^2 + \lambda_r \log(p(\mathbf{x})) \quad (2)$$

The regularization process is performed in the dual-tree complex wavelet transform (DT-CWT) domain as it provides near shift-invariance and increased directional selectivity compared to conventional wavelet transforms [5]. The wavelet transform of an image tends to be sparse which eases the formulation of expectations. We denote $(\mathbf{W}_j)_{j \in S}$ and $(\mathbf{M}_j)_{j \in S}$ as the decomposition and reconstruction wavelet subbands, respectively, where $S = 0, 1, \dots, S_t$ and S_t is the total number of wavelet subbands for all decomposition levels. We also assume that the real and imaginary parts of the transform's outputs are treated as separate coefficients so that \mathbf{W}_j and \mathbf{M}_j are real matrices thereby producing real valued $\mathbf{w}_j = \mathbf{W}_j\mathbf{x}$ and $\mathbf{m}_j = \mathbf{M}_j\mathbf{w}_j$ for a real image \mathbf{x} .

For the prior knowledge $\log(p(\mathbf{x}))$, the magnitudes of highpass coefficients of wavelets can be employed, since they reflect image features, such as lines, corners and textures.



Fig. 1. Top: Original ground plane. Bottom: Rectified ground plane.

Here, $\log(p(\mathbf{x}))$ is thus set to be $\mathbf{w}^T \mathbf{\Gamma} \mathbf{w}$, where $\mathbf{\Gamma}$ is a weight matrix. This can be seen as the weighted L2-norm which can be utilised to approximate the L0 sparseness measure by defining $\gamma_j = 1/(|w_j|^2 + \epsilon)$, $\gamma_j \in \mathbf{\Gamma}$, where ϵ is a small number for preventing zero denominator. Employing the subband pre-emphasis process from [3], the cost function is finally obtained as shown in Eq. 3.

$$\begin{aligned} \bar{C}(\mathbf{w}) = & \|\mathbf{y} - \mathbf{A}\mathbf{H}\mathbf{B}\mathbf{x}\|^2 + \lambda_r \mathbf{w}^T \mathbf{\Gamma} \mathbf{w} \\ & + \sum_{j \in S} \alpha_j \|\mathbf{W}_j \mathbf{x} - \mathbf{w}_j\|^2 - \|\mathbf{A}\mathbf{H}\mathbf{B}\mathbf{x} - \mathbf{A}\mathbf{H}\mathbf{B}\mathbf{M}\mathbf{w}\|^2 \end{aligned} \quad (3)$$

2.1. Restoration Process

We divide the observed image $\bar{\mathbf{y}}$ vertically into K overlapped regions according to their distance from the camera, which is taken as from the bottom part of the observed image. Then, each region $\bar{\mathbf{y}}_k, k = 0, 1, \dots, K-1$ is processed individually. The estimated projective region $\bar{\mathbf{x}}_k$ is achieved by minimising Eq. 3. That is $\partial \bar{C}(\mathbf{w}_k, \forall j \in S) / \partial \mathbf{w}_k = 0$. Then, the iterative process is operated as Eq. 4 and 5.

$$\mathbf{z}_j = \mathbf{W}_j \left(\alpha_j \mathbf{x}_k^{(n)} + \mathbf{B}^T (\mathbf{A}\mathbf{H})^T (\bar{\mathbf{y}}_k - \mathbf{A}\mathbf{H}\mathbf{B}\mathbf{x}_k^{(n)}) \right) \quad (4)$$

We generate the initial reprojective image $\mathbf{x}_k^{(0)} = (\mathbf{A}\mathbf{H})^T \bar{\mathbf{y}}_k$ using a known \mathbf{H} with bicubic interpolation. Following the method in [3], a weight ω_j is created using structure obtained from wavelet coefficients and estimated noise variance σ^2 . The weight for each subband is defined as $\omega_j = 1 / (\alpha_j + \lambda_r \sigma^2 \gamma_j)$.

$$\mathbf{x}_k^{(n+1)} = \sum_{j \in S} \mathbf{M}_j (\omega_j \mathbf{z}_j) \quad (5)$$

where $\mathbf{x}_k^{(n+1)}$ and $\mathbf{x}_k^{(n)}$ are the estimated projective regions at the $(n+1)$ th and n th iterations. α_j is a relaxation parameter controlling rate of convergence (Here $\alpha_j = 1$).

The final estimated projective region $\bar{\mathbf{x}}_k$ is achieved when the difference between $\mathbf{x}_k^{(n+1)}$ and $\mathbf{x}_k^{(n)}$ is less than a threshold. Finally the estimated projective image $\bar{\mathbf{x}}$ is constructed by combining all regions as shown in Eq. 6, where \mathbf{J} is a matrix of ones with the same size as $\bar{\mathbf{x}}$ and \mathbf{a}_k is a weight mask of the region k .

$$\bar{\mathbf{x}} = \sum_{k=0}^{K-1} \mathbf{a}_k \bar{\mathbf{x}}_k \quad (6a)$$

$$\mathbf{J} = \sum_{k=0}^{K-1} \mathbf{a}_k \quad (6b)$$

2.2. Blur Kernel Estimation

We assume that the blur kernel corresponding to each area k is space-invariant so that the deconvolution process can be performed using a single point-spread function (PSF). This means, $\mathbf{B}_k \mathbf{x}_k$ is a discrete convolution of \mathbf{x}_k with the PSF and

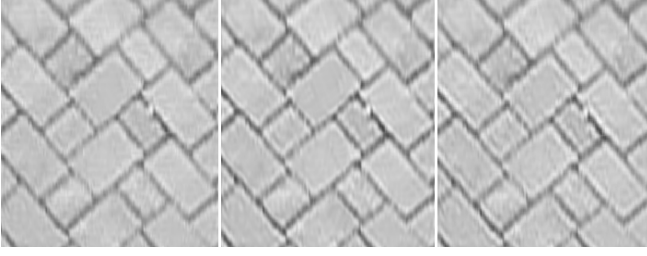


Fig. 2. Restored images with different sizes of blur kernels. Left to Right: $s_k/2$, s_k and $2s_k$, respectively

can be computed by multiplying together the discrete Fourier transform (DFT) coefficients of \mathbf{x}_k and the DFT coefficients of the PSF, then taking the inverse DFT of the product.

We also assume that the closest area to the camera, is as sharp as the original projective image; therefore, it is representative of the sharpness required. To estimate \mathbf{B}_k , we generate the estimated observed area $\tilde{\mathbf{y}}_k$ using information from area 0. That is, $\tilde{\mathbf{y}}_k = \mathbf{A}\mathbf{H}_k\tilde{\mathbf{x}}_0$. Subsequently, the blur kernel can be estimated using Eq. 7.

$$\tilde{\mathbf{x}}_0 \approx (\mathbf{A}\mathbf{H}_0)^T \mathbf{y}_0 \quad (7a)$$

$$\mathbf{b}_k = \Theta \left(\frac{\text{DFT}((\mathbf{A}\mathbf{H}_k)^T \mathbf{A}\mathbf{H}_k \tilde{\mathbf{x}}_0)}{\text{DFT}(\tilde{\mathbf{x}}_0)}, \mathbf{s}_k \right) \quad (7b)$$

$$\tilde{\mathbf{b}}_k = \mathbf{b}_k / \max(\mathbf{b}_k) \quad (7c)$$

where $\Theta(\mathbf{O}, \mathbf{s})$ converts the optical transfer function array \mathbf{O} , to a point-spread function array of specified size \mathbf{s} using the inverse DFT, centred at the origin. Finally the blur matrix for each region \mathbf{B}_k is obtained from $\Theta^T(\tilde{\mathbf{b}}_k, M \times N)$, where Θ^T computes the DFT of $\tilde{\mathbf{b}}_k$ with the same size as \mathbf{x}_k , which is $M \times N$.

The kernel size is computed by transforming four positions in current area, namely the centre position of $\tilde{\mathbf{y}}_k$, (p_i, p_j) and its three neighbours, $(p_i + 1, p_j)$, $(p_i, p_j + 1)$ and $(p_i + 1, p_j + 1)$, to the corresponding positions on projective area, namely (q_{i1}, q_{j1}) , (q_{i2}, q_{j2}) , (q_{i3}, q_{j3}) and (q_{i4}, q_{j4}) , respectively. Then, $\mathbf{s}_k = (s_i, s_j)$ is obtained from Eq. 8 and 9.

$$s_i = \lceil (|q_{i2} - q_{i1}| + |q_{i4} - q_{i3}|) / 2 \rceil \cdot 2 + 1 \quad (8)$$

$$s_j = \lceil (|q_{j3} - q_{j1}| + |q_{j4} - q_{j2}|) / 2 \rceil \cdot 2 + 1 \quad (9)$$

Fig. 2 shows the effect of the size of \mathbf{b}_k . If the kernel size is too small, the restored result will still be blurred. In contrast, if the kernel size is too large, kernel-related artifacts become visible.

3. RESULTS AND DISCUSSION

Four example ground surface images of bricks, grass, sand and tarmac are shown in Fig. 3. These images have a res-



Fig. 3. Top-left: *bricks*. Top-right: *grass*. Bottom-left: *sand*. Bottom-right: *tarmac*.

olution of 1920×1080 pixels with a 24-bit RGB colour format. Four base points on the ground, marked as a $2\text{m} \times 2\text{m}$ square were used for computing the homography \mathbf{H} . Based on the anti-aliasing matrix \mathbf{A} , the interpolation kernel size is extended according to the down-sampling scale, resulting in increased smoothness in the shrinking area. Also, in the work reported here, we set parameters as follows: $\sigma^2 = 10^{-5}$ (approximately 40 dB noise) and the regularization parameter, $\lambda_r = 1$. For the iterative process, the convergence threshold was set to 10^{-5} which resulted in fewer than 15 iterations to achieve satisfactory results.

Subjective results for our new approach are illustrated in Fig. 4. These compare the projective images obtained from traditional homography (using bicubic interpolation) with those obtained from our proposed projective image restoration method. The figure shows the sand and bricks images cropped at 4m, the grass image at 6m and the tarmac image at 7.5m. The enhanced sharpness resulting from the proposed method is clearly visible in these subjective comparisons.

Finally, we investigate the performance of a surface-type classifier based on our restored images compared to that using images constructed by bicubic interpolation. We employ the 4 ground types discussed above: bricks, sand, tarmac and grass, for classification and we divide each image into 6 distant ranges: $<2\text{m}$, 2.5m , 5m , 7.5m , 10m and $>10\text{m}$. At each range, texture features are employed as training data in the classification process (a list of texture features can be found in [6]). Texture features at other distances are used for evaluating system performance. The aim of this study is to investigate whether our proposed restoration achieves an improvement in surface-type recognition at long range.

LIBSVM for MATLAB [7] was employed for multiclass classification (4 classes were used in this study). The classification accuracy using the conventional projective images together with those obtained with our proposed restoration method are shown in Tables 1 and 2, respectively. These tables show the training data range across each row with the classification results for different distances in each column.



Fig. 4. Projective images. Left: Traditional homography. Right: Proposed restoration

Table 1. Classification accuracy using projective images by bicubic interpolation

Range	Accuracy %						Avg
	<2m	2.5m	5m	7.5m	10m	>10m	
<2m	100.0	95.5	78.6	69.8	52.9	35.0	72.0
2.5m	100.0	100.0	94.8	73.5	44.8	25.0	73.0
5m	79.6	93.3	100.00	94.1	69.1	38.0	79.0
7.5m	68.7	77.9	95.5	100.0	100.0	87.0	88.2
10m	34.3	60.2	98.5	100.0	100.0	100.0	82.2
>10m	40.6	64.7	75.0	90.4	100.0	100.0	78.4
Avg	70.5	81.9	90.4	87.9	77.8	64.1	78.8

The results clearly show improved performance for the projective images generated by the proposed method, with an improvement in overall classification accuracy by approximately 5.5%. The best performance (with an average accuracy of 89.2%) is achieved when using data at a range of 2.5m for training.

Table 2. Classification accuracy using projective images by proposed image restoration

Range	Accuracy %						Avg
	<2m	2.5m	5m	7.5m	10m	>10m	
<2m	100.0	100.0	85.3	72.1	56.3	58.8	78.8
2.5m	100.0	100.0	94.12	88.2	77.21	75.6	89.2
5m	81.9	96.9	100.0	97.1	80.0	70.6	87.7
7.5m	64.7	78.4	96.0	100.0	100.0	88.2	87.9
10m	50.0	61.5	89.1	100.0	100.0	100.0	83.4
>10m	50.0	60.0	71.0	91.3	100.00	100.0	78.7
Avg	74.4	82.8	89.3	89.6	85.3	84.3	84.3

4. CONCLUSIONS

This paper presents a new method for restoring projective (rectified) surface images in order to reveal increased detail at points far from the camera plane. Such detail is not apparent in the original image obtained through conventional acquisition. The proposed restoration process employs wavelet-based deconvolution with L0 reweighted-L2 minimisation. As projective transform and interpolation introduce range-dependent blurring, we propose a means of blur kernel estimation based on characteristics contained in the original image, biased according to the relative position of the target region in the projective image. Subjective comparisons show a clear improvement associated with the proposed scheme over the conventional projective transform and bicubic interpolation. Furthermore surface-type classification (using an SVM classifier) is improved by approximately 5% compared to the case when conventional homography is employed.

5. REFERENCES

- [1] Edmund Y. Lam and Joseph W. Goodman, "Iterative statistical approach to blind image deconvolution," *J. Opt. Soc. Am. A*, vol. 17, no. 7, pp. 1177–1184, July 2000.
- [2] M. Hirsch, S. Sra, B. Scholkopf, and S. Harmeling, "Efficient filter flow for space-variant multiframe blind deconvolution," in *Computer Vision and Pattern Recognition (CVPR)*, June 2010, pp. 607–614.
- [3] Yingsong Zhang and N. Kingsbury, "Restoration of images and 3d data to higher resolution by deconvolution with sparsity regularization," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, Sept. 2010, pp. 1685–1688.
- [4] Anestis Antoniadis and Jianqing Fan, "Regularization of wavelet approximations," *Journal of the American Statistical Association*, vol. 96, no. 455, pp. 939–967, 2001.
- [5] I.W. Selesnick, R.G. Baraniuk, and N.G. Kingsbury, "The dual-tree complex wavelet transform," *Signal Processing Magazine, IEEE*, vol. 22, no. 6, pp. 123–151, Nov. 2005.
- [6] N. Anantrasrichai, Alin Achim, James E Morgan, Irina Erchova, and Lindsay Nicholson, "Svm-based texture classification in optical coherence tomography," in *International Symposium on Biomedical Imaging*, 2013.
- [7] C.C. Chang and C.J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.