Tapping Motion Blur for Robust Normal Estimation of Planar Scenes

Subeesh Vasu, [1], A.N. Rajagopalan[1] and Gunasekaran Seetharaman[2]

[1]Department of Electrical Engineering, Indian Institute of Technology Madras.
[2]Information Directorate AFRL/RIEA, Rome NY, USA

This is a draft version of the paper accepted for presentation at International Conference on Image Processing (ICIP) SEPTEMBER 2015. The final copyrighted version is available at
http://ieeexplore.ieee.org/document/7351305/

TAPPING MOTION BLUR FOR ROBUST NORMAL ESTIMATION OF PLANAR SCENES

Subeesh Vasu, A.N. Rajagopalan

Department of Electrical Engineering
IIT Madras, Chennai, India
subeeshvasu@gmail.com, raju@ee.iitm.ac.in

Gunasekaran Seetharaman

Information Directorate
AFRL/RIEA, Rome NY, USA
guna@csc.lsu.edu

ABSTRACT

We propose a framework for robust estimation of normal of a planar scene from a single motion blurred observation. We first reveal how feature points can be extracted from blur kernels and matched to generate several point correspondences. Although these points correspond to different homographies, the fact that they conform to the same normal yields a rank-3 constraint which we harness within a hierarchical clustering framework to estimate the normal accurately.

Index Terms— Motion blur, plane normal, blur kernel, feature points, correspondence, homography, clustering.

1. INTRODUCTION

Many works exist in literature that specifically address the task of inferring *planar* scene geometry. Not only do planar scenes lend themselves to elegant representation but also provide useful interpretation for performing high-level scene description, classification and recognition [1, 2, 3]. Among the works that estimate relative orientation of planar scenes, [4] exploits foreshortening of the texture auto-correlation function of an oriented surface by assuming the surface texture to be isotropic. In [5], spatial variation of a single dominant local frequency component of the image texture is used to estimate orientation. Farid et al. [6] have proposed a method based on the observation that higher-order correlations in frequency domain caused by perspective projection of the scene are proportional to the orientation of the plane. Horizontal and vertical vanishing points extracted from the image of a text plane are used in [7]. McCloskey et al. [8] estimate planar orientation from a single image using optical blur as a cue. Recently, motion blur is used as a cue in [9] for normal estimation by using the extremities of blur kernels.

We propose a method for normal estimation from a single motion blurred observation of a planar scene. While we too assume translational blur as in [9], there are important differences. Unlike [9] that employs only the extreme points, we propose a systematic method to extract reliable features from blur kernels and match them across space-varying blur kernels within a single blurred image. Due to the inherent ambiguity in the blur kernel estimation process, the extreme points often tend to be quite noisy and affect the accuracy of the estimated normal. In contrast, our ability to match multiple feature points across space-varying blur kernels enables us to exploit the constraint that multiple homographies must conform to the same normal. In our work, this is harnessed within a hierarchical clustering framework to yield a robust solution. Although it is known that feature points cannot be matched across motion blurred images, our work interestingly reveals that within a single blurred image it is possible to match feature points of blur kernels extracted from different spatial locations. This opens up several possibilities beyond normal estimation including potential usage in non-fronto parallel deblurring, change detection in differently oriented images, and image splicing under arbitrary orientation, to name a few. Our main contributions are summarized below.

• This work improves upon existing works by imposing a judicious condition that 'multiple homographies should conform to a single normal' yielding a rank-3 constraint for estimation of surface orientation.

• Unlike [9] that relies only on extreme points, we advocate use of multiple feature points within a hierarchical clustering framework for achieving robustness.

• Although it is common to extract and match feature points across images, this is the first work of its kind to unveil that it is possible to systematically extract and match feature points across *blur kernels*.

2. KERNELS AND CORRESPONDENCES

The setup considered in this work is as shown in Fig. 1. It is well-known that for planar geometry, pixel displacements corresponding to different camera positions can be related through a global homography. Because we restrict the camera motion to be in-plane translations, the homography H_ν which relates the pixel coordinates of the reference frame with those of the image captured by the camera at time ν, will have the form $H_\nu = K\left(I + t_\nu \frac{n^T}{d}\right)K^{-1}$. Here I is 3×3 identity matrix, $t_\nu = [T_{X_\nu}\ T_{Y_\nu}\ 0]$ is the translation vector at time ν, d is the perpendicular distance from the camera center to the plane, $n = [N_X\ N_Y\ N_Z]^T$ is the surface normal of the plane, and K is the camera intrinsic matrix with focal length f.

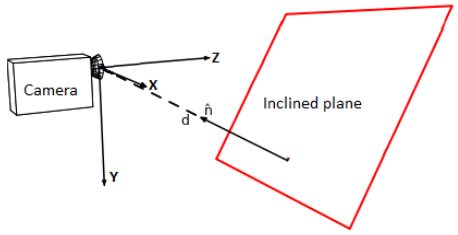
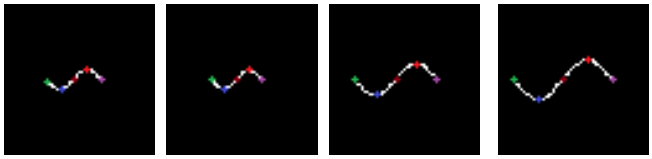


Fig. 1: Camera setup.

Because of relative motion between camera and scene during the exposure time T_e, light intensities from multiple scene points get averaged at the camera sensor resulting in a motion blurred image. The blur itself can be space-invariant or otherwise, depending on camera motion and relative orientation of the plane. The blur kernel at any location $\mathbf{x}$ can be expressed as a function of the set of homographies induced by camera motion [10], [11] as

$$h(\mathbf{x}, \mathbf{s}) = \frac{1}{T_e} \int_0^{T_e} \delta\left(\mathbf{s} - (H_\nu(\mathbf{x}) - \mathbf{x})\right) d\nu \qquad (1)$$



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

Fig. 2: Blur kernel for (a) fronto-parallel scene, and (b-d) inclined plane for different positions on the blurred image.

Fig. 2(a) shows an example blur kernel at a random location in a blurred image. Different points on the blur kernel are shown with different colors and they are influenced by different homographies as governed by equation (1). For a space invariant situation (such as a fronto-parallel scene) the blur kernel will be identical everywhere. On the other hand, for a space variant scenario such as an inclined plane, the blur kernel varies as a function of the spatial location in the image as shown in Figs. 2(b-d). The points that are governed by the same homographies i.e. point correspondences, are shown in the same color across the blur kernels. Although different points on a single blur kernel correspond to different homographies, note that the underlying normal is the same. Thus, for each of these homographies, if their point correspondences can be determined, they can serve as a valuable cue for estimating the underlying normal of the plane.

3. FEATURE DETECTION IN BLUR

Let (x_i, y_i) refer to the spatial location on the blurred observation from where the i^{th} blur kernel is estimated. Let (x_{ν_i}, y_{ν_i}) denote the x and y coordinates corresponding to (x_i, y_i) with respect to camera translation t_ν. It can be shown [9] that these correspondences can be expressed using the homography relation in equation (1) as

$$\begin{bmatrix} x_{\nu_1} \\ x_{\nu_2} \\ . \\ . \\ x_{\nu_\alpha} \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ . & . & . \\ . & . & . \\ x_\alpha & y_\alpha & 1 \end{bmatrix} \begin{bmatrix} 1 + N_X \frac{T_{X_\nu}}{d} \\ N_Y \frac{T_{X_\nu}}{d} \\ f N_Z \frac{T_{X_\nu}}{d} \end{bmatrix} \qquad (2)$$

where $i = 1, 2, .., \alpha$, and α is the number of blur kernels. A similar set of equations can be written for the y-coordinate too. From equation (2) it is clear that the unknown normal can be estimated provided $\alpha \geq 3$ i.e. the correspondences are known for a point across 3 or more blur kernels.

The method in [9] uses extreme points on the blur kernels to estimate the normal. But there are associated issues. Since the blur kernels are estimated independently, kernels returned by blind deblurring methods such as [12] often have different and unknown spatial shifts because multiple choices of the focused image and point spread function (PSF) can give rise to the same blurred image. The authors at [9] attempt to handle this problem by using differences of the spatial location of the extreme points. This is because change in coordinate positions for any pair of points is insensitive to the shift in the blur kernel as a whole. However, this does not solve the issues entirely. Deblurring being an inverse problem, it can introduce distortions in the estimated blur kernels making extreme points unreliable. In [9], the extreme points are detected by thresholding the blur kernel followed by row-column sum. However, since binarization is sensitive to thresholding, one may not always detect the true end point. In this case the argument in [9] that taking the difference of end points will nullify the unknown shift in the kernel is no longer valid. This can also occur when end-point detection using the row-column sum approach goes wrong due to the presence of spurious points around the kernel. Presence of even a few false correspondences (outliers) can cause a large deviation in the estimated normal. This motivated us to explore the possibility of using correspondences from multiple feature points and not just the extremities.

There are many feature point detectors such as SIFT [13], SURF [14] , Hessian Laplace, Harris Laplace (HL) [15] etc. which have exhibited excellent performance for feature extraction in different applications. While HL gives corner points as features, other methods are blob detectors. We observed that most of the feature points given by blob detectors such as SIFT are located outside the blur kernel and are therefore not meaningful for our application. On the other hand, HL is guaranteed to give almost all the feature points within the blur kernel. Figs. 3(a) and 3(b) show features detected on a blur kernel using SIFT and HL detectors, respectively. The radius of the circle around the feature points is proportional to the scale at which the feature points are detected.

(a) (b)

Fig. 3: Feature points detected by (a) SIFT and (b) HL .

For matching, feature descriptors for the extracted feature points are determined using the SIFT descriptor [13]. Feature matching across the blur kernels is performed by finding the nearest neighbor which satisfies the ratio test. To compare overall performance, we found the percentage of feature points with correct match [13] for both the methods across randomly selected pairs of blur kernels. The fraction of feature points with correct match was observed to be about .83 and .92 for SIFT and HL, respectively. Due to the above reasons, we chose to employ Harris Laplace for feature detection. Feature points, if extracted directly from the kernel, are typically few in number. Hence, we first interpolate the blur kernels using cubic spline as this yields a more natural and smoother extension of the kernels. The difference in spatial location $\triangle x_{p_i}$ of a pair of points on the i^{th} kernel yields

$$\begin{bmatrix} \triangle x_{p_1} \\ \triangle x_{p_2} \\ . \\ . \\ \triangle x_{p_\beta} \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ . & . & 1 \\ . & . & 1 \\ x_\beta & y_\beta & 1 \end{bmatrix} \begin{bmatrix} N_X \frac{\triangle T_{X_\nu}}{d} \\ N_Y \frac{\triangle T_{X_\nu}}{d} \\ fN_Z \frac{\triangle T_{X_\nu}}{d} \end{bmatrix} \quad (3)$$

where β is number of kernels across which correspondences have been found for the above pair of points and $\triangle T_{X_\nu}$ is translational difference. Equation (3) can be solved to yield an estimate of the normal. Although [9] uses only extreme points, we could potentially employ any pair of points.

4. ROBUST NORMAL ESTIMATION

Since normal estimated from a single feature pair may not be reliable due to outliers, we propose to solve equation (3) for every possible pair of feature points. Although each pair follows a different homography, they should all ideally return the same normal. However, this seldom happens in practice due to wrong correspondences and it is difficult to arrive at the true normal. While the notion of a common normal underlying the homographies is theoretically sound, the challenge is to develop a framework that can incorporate this condition. We next bring into bearing the following important constraint.

Claim: *The set of all homographies of a planar scene induced by in-plane translational motion of the camera lies in a 3-dimensional space.*

Proof: Homography H_ν corresponding to in-plane translation of the camera and a planar scene can be obtained by expanding the basic homography equation as

$$H_\nu = \begin{bmatrix} 1 + N_X \frac{T_{X_\nu}}{d} & N_Y \frac{T_{X_\nu}}{d} & fN_Z \frac{T_{X_\nu}}{d} \\ N_Y \frac{T_{Y_\nu}}{d} & 1 + N_Y \frac{T_{Y_\nu}}{d} & fN_Z \frac{T_{Y_\nu}}{d} \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

The column matrix h_ν obtained by lexicographically arranging elements of H_ν can be decomposed as

$$h_\nu = N \begin{bmatrix} fT_{X_\nu} & fT_{Y_\nu} & 1 \end{bmatrix}^T \quad (5)$$

where

$$N^T = \begin{bmatrix} \frac{N_X}{fd} & \frac{N_Y}{fd} & \frac{N_Z}{d} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{N_X}{fd} & \frac{N_Y}{fd} & \frac{N_Z}{d} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Note that N depends on the normal of the plane and not on the camera motion and hence remains the same for all the homographies. On the other hand $s_\nu = [fT_{X_\nu}\ fT_{Y_\nu}\ 1]^T$ depends only on camera motion and will be different for different homographies. A collection of p homographies of a planar scene under in-plane translations can be expressed as

$$[h_{\nu 1}\ h_{\nu 2} \ldots h_{\nu p}] = N[s_{\nu 1}\ s_{\nu 2} \ldots s_{\nu p}] \quad (6)$$

As N is a matrix of rank utmost 3, the matrix obtained by stacking p homographies is also of rank utmost 3.

We choose the largest blur kernel as reference and find matches for all the feature points in it, across other blur kernels. In order to reduce the computational complexity and to improve accuracy, we use (in equation (3)) only those feature pairs which are located a minimum distance apart. However, due to the sources of errors mentioned in Section 3, it is difficult to ascertain which one of these normals is correct. To solve this issue, we resort to a clustering-based framework. Since naive k-means is sensitive to initial estimates, we adopt hierarchical clustering to group the normals. We build a hierarchical cluster tree using an average linkage clustering algorithm [16], with distance measure being the angle between normals. We chose average linkage as it was found to outperform single linkage, complete linkage etc.

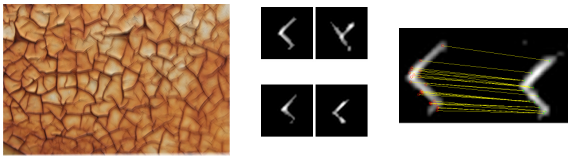
The projection error is found out for each cluster using the homographies corresponding to that cluster. While the normal estimates obtained using equation (3) are derived by assuming in-plane translations, we need to find general homography from the correspondences and employ the rank-3 constraint to arrive at the desired normal. Towards this end, the projection error of a collection of homographies is computed as the residual error [17] corresponding to the 3-rank approximation of the matrix formed by stacking lexicographically arranged homographies, using singular value decomposition (SVD). The mean of the cluster with the lowest projection error is treated as the final estimate for the normal. The proposed method is summarized under Algorithm 1.

Algorithm 1 : Robust normal estimation of a plane

1. Select N points on the blurred image at random. Estimate PSFs of patches around these points using [12].
2. Spline-interpolate and select largest kernel as reference.
3. Extract feature points from all the blur kernels using Harris Laplace and get their SIFT descriptors.
4. Select all feature pairs in the reference kernel which are apart by a minimum distance and get their correspondences.
5. Estimate the normal from each pair using equation (3).
6. Apply hierarchical clustering on all the normals and find rank-3 approximation of the homographies in each cluster.
7. The mean of the cluster with the lowest projection error yields the desired normal.

5. EXPERIMENTAL RESULTS

Since estimation of general homographies requires atleast 8 point correspondences, we randomly extract 14 blur kernels from the blurred image by using patches of size 120×120. The blur kernels were interpolated by a factor of 2 prior to feature detection. On an average, about 30 feature points could be matched. For real examples, we used a Canon 60D camera and the focal length from meta data was used for normal estimation. Since the source code was available only for [9] (provided by the authors), we compare our performance with that method alone. All results are best viewed in pdf.



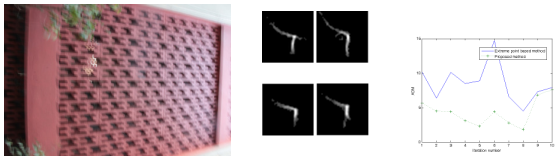
(a) (b) (c)

Fig. 4: (a) Synthetic example. (b) PSFs at different locations of (a). (c) Depiction of matches between two kernels.

Fig. 4(a) shows a synthetically blurred image with known normal $n = [0.5\ 0\ 0.866]$. From Fig. 4(b), observe that the extracted blur kernels are quite challenging to deal with. The estimated normal using [9] was found to be [0.8592 0.2793 0.4286] which deviates from the true normal by more than 15 degrees. Using our method, the normal was found to be [0.4618 0.0813 0.8832] which is in error by only 5.2 degrees.

For fronto-parallel real scenes (i.e. n = [0 0 1]), we could easily confirm the correctness of the estimated normal. For unknown inclination, we validate the accuracy by comparing the estimates of normal obtained using different collections of blur kernels from the same blurred image. We compute the average angular deviation from mean (AADM) to asses the accuracy of the estimated normal.

For the real example in Fig. 5(a), the mean of the estimated normals obtained using our method was found to be



(a) (b) (c)

Fig. 5: (a) Inclined wall with normal oriented towards X-axis. (b) Extracted PSFs. (c) ADM for both methods.

[0.3885 -0.0048 0.9214] while the extreme point method gave a mean of [0.2676 -0.1231 0.9556]. The AADM was 8.5 degrees for [9] and 4.3 degrees for our method. Angular deviation from mean (ADM) for both the methods and for 10 different collection of blur kernels are shown in Fig. 5(c).

(a) (b) (c)

Fig. 6: (a) Inclined ground plane with normal oriented towards Y axis. (b) Extracted PSFs. (c) ADM for (a).

Fig. 6(a) shows a real example of an inclined ground plane. The mean of the estimated normals was found to be [0.1325 0.7622 0.6336] and [0.0189 0.7111 0.7028], using our method and [9], respectively. The AADM values were 12.2 degress for [9] and 4.5 degrees for our scheme. Due to space constraints, we have given only representative results here. However, we have verified that the proposed method comfortably outperforms [9] in challenging conditions. The unoptimized version of our code takes about 6 seconds to execute in comparison to 2.6 seconds for [9].

6. CONCLUSIONS

We proposed a robust normal estimation framework for planar scenes that uses multiple feature correspondences across blur kernels. Although the correspondences are governed by different homographies we exploit the fact that they all conform to the same normal and employ a rank-3 constraint within a hierarchical clustering framework to estimate the underlying normal. We plan to extend the scope to multiple planes.

Acknowledgement: Partial support through a grant from the Asian Office of Aerospace Research and Development, AOARD/AFOSR is gratefully acknowledged. The results and interpretations presented in this paper are that of the authors, and do not necessarily reflect the views or priorities of the sponsor, or the US Air Force Research Laboratory.

7. REFERENCES

[1] D. Hoiem, A. A. Efros, and M. Hebert, "Geometric context from a single image," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. IEEE, 2005, vol. 1, pp. 654–661.

[2] D. Hoiem, A. A. Efros, and M. Hebert, "Closing the loop in scene interpretation," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[3] D. Hoiem, A. N. Stein, A. A. Efros, and M. Hebert, "Recovering occlusion boundaries from a single image," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[4] L. G. Brown and H. Shvaytser, "Surface orientation from projective foreshortening of isotropic texture autocorrelation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 6, pp. 584–588, 1990.

[5] B. J. Super and A. C. Bovik, "Planar surface orientation from texture spatial frequencies," *Pattern Recognition*, vol. 28, no. 5, pp. 729–743, 1995.

[6] H. Farid and J. Kosecka, "Estimating planar surface orientation using bispectral analysis," *Image Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 2154–2160, 2007.

[7] P. Clark and M. Mirmehdi, "Estimating the orientation and recovery of text planes in a single image.," in *BMVC*. Citeseer, 2001, pp. 1–10.

[8] S. McCloskey and M. Langer, "Planar orientation from blur gradients in a single image," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 2318–2325.

[9] M. P. Rao, A.N. Rajagopalan, and G. Seetharaman, "Inferring plane orientation from a single motion blurred image," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, 2014, pp. 2089–2094.

[10] M. Sorel and J. Flusser, "Space-variant restoration of images degraded by camera motion blur," *Image Processing, IEEE Transactions on*, vol. 17, no. 2, pp. 105–116, 2008.

[11] C. Paramanand and A.N. Rajagopalan, "Shape from sharp and motion-blurred image pair," *International Journal of Computer Vision*, vol. 107, no. 3, pp. 272–292, 2014.

[12] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *Computer Vision–ECCV 2010*, pp. 157–170. Springer, 2010.

[13] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[14] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer Vision–ECCV 2006*, pp. 404–417. Springer, 2006.

[15] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Computer VisionECCV 2002*, pp. 128–142. Springer, 2002.

[16] A. D. Gordon, "A review of hierarchical classification," *Journal of the Royal Statistical Society. Series A (General)*, pp. 119–137, 1987.

[17] L. Zeinik-Manor and M. Irani, "Multiview constraints on homographies," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 2, pp. 214–223, 2002.