CSE 559A : Forward Motion Deblurring

Aravindh Shanmuganathan
shanmuganathan.a@wustl.edu

Abstract

For my final project, I selected the paper presented in 2013 ICCV journal on Forward Motion deblurring.This paper provides a solution for a special type of blurring caused due to motion in the forward direction, known as out-of-plane translation. This problem has several applications such as in dash mounted cameras and traffic cameras. The paper provides a model to represent this problem and an iterative estimation algorithm to estimate the correct values of weights and sharp images. Homography is used to relate the image captured at different camera positions. I have partially implemented this paper and analyzed its limitations and advantages over previous works.

Figure 1: Motion blurring due to out of plane translation

1 Introduction

Motion blur is a problem that is faced in a variety of applications where either the camera is in motion or the scene is not static. There have been several different approaches to tackling this problem. But still due to the nature in which motion blur is caused, varies in each case there has been no algorithm which fits all problems. This is mainly due to the several different types of motion blur which could occur in several possible combinations of rotation or translation about the 3 axes. Due to the fact that any camera has 6 degrees of freedom in the 3D world, it is very difficult to find a general model that accurately represents the different kinds of motion.

The reason we are studying this problem is because of the enormous amount of images that are distorted due to some kind of shake during the capture of the photo. Some examples include capturing photos while walking or in a car. Other kind of examples include dynamic scene examples where the environment is in motion such as wildlife photography, sports photography,etc.

This paper separates itself from the previous works by addressing the specific problem of images blurred due to motion in the forward or backward direction along the z-axis, or out of plane translation. Although there have been papers which address the deblurring problem, they work well only under their set of specified constraints which often restricts motion along the z-axis. This paper provides a solution for several applications such as deblurring images captured by traffic cams and dashboard mounted cameras inside vehicles.The paper also speaks about the algorithm being used in the identifying license plates and road signs.

These applications are very well suited for this algorithm due to the way that it is modelled. In any scene involving forward motion, most valuable information is present on flat surfaces at certain angles to the camera sensor plane, such as road information boards and license plates, and human faces which can be regarded as planar surfaces from a certain distance. Therefore this paper is successful due to eliminating addressing images that are blurred due to camera rotation and only focusing on deblurring the 3D planes in any image with forward motion blur.

2 Background & Related Work

In some of the previous works discussed in this paper, there are several methods for deblurring splitting them into spatially variant and spatially invariant models. Spatially invariant blurring refers to the case where the blurred image has uniform blurring meaning that all pixels can be correspond to a single kernal, i.e. each blurred image is formed by a single sharp image and a single kernal. Whereas for spatially variant models or non-uniform blind deconvolution the image is not uniformly blurred meaning the image may be more blurred to greater extents and in different ways than other parts of the image. This type of model involves convolving a sharp image with a union of different kernals to produce the single blurred image. Naturally since the non-uniform deconvolution assumes a more generalized model for blurring it tends to have more applications and match the real world instances of blurred images.

Some of the notable work done in this field that the author called back on in the field of uniform deconvolution, there was Fergus et al. [1] who used a blind deconvolution on natural images using a Bayesian approach to estimate the blur kernal through marginal probability maximization. Another paper which seemed to employ a unique method was Cho and Lee [2] who used edge prediction to estimate and optimize the kernal matrix. The author also goes over several other methods some of which involved optimizing on the above methods or another by Xu and Jia [3] which identified structures which might be detrimental to the estimation of the kernal.

In the case of non-uniform deconvolution, there were methods that addressed blurring caused by rotation about a certain axis. The first paper to address this was Shan et al [4], who only considered in plane rotation. The author also references Tai et al [5] who used projective models to handle spatially variant blur.

What is importantly noticed was that in most previous models the depth of the image or movement in the z axis is fixed as a constant constraint, only in one paper by Xu and Jia [6], they used a pair of stereo images to estimate the sharp image. Thus the work of this paper in estimating the sharp image from a single blurred image is a significant step in the work of deblurring methods.

2.1 Background

The paper has explained the general approach assumed by most methods and the general model that they work from. This can be summarized as , in every scene X where and image is captured onto the sensor plane , every element in the image whose homogenous coordinates can represented as x = $(x,y,1)^T$ and the scene X = $(x,y,z,1)^T$. In most methods since the motion along depth i.e. z axis is considered constant the model of the blurred image is considered to be a sum of the sharp image multiplied with a homography matrix that relates the images between the different instances of the scene or camera motion and a weight value. The weight value corresponds to the duration of each sharp image in a certain pose. This is added with a certain amount of noise. The following equation represents the described blur model:

$$b = \sum_i w_i P_i l + \epsilon \tag{1}$$

where b is the blurred image, w_i is the weight value or duration of each sharp image in a certain pose, P_i is the N*N transformation matrix that transforms the original position sharp image l to the image in the current position sharp image l_i. And ϵ is the additive noise.

Here the transformation between l and l_i is represented by the projection matrix, by tow view geometry any pixel x in l is related to the corresponding pixel x' in another pose l_i as follows:

$$x' = KRK^{-1} + Kt/z = Hx \tag{2}$$

Where R represents the Rotation matrix and t represents the translation vector about the three axes. K is the intrinsic camera matrix, which remains fixed due to a constant focal length. The entire transformation matrix H can be represented as:

$$H = K \begin{pmatrix} 1 & -\theta_z & \theta_y + t_x/z \\ \theta_z & 1 & -\theta_x + t_y/z \\ -\theta_y & \theta_x & 1 + t_z/z \end{pmatrix} K^{-1} \tag{3}$$

Now that we have defined the general model in which blurred images are represented, the weight value or duration of each sharp image in a certain pose must be estimated. The vector of weights $w = (w_1, w_2, ..., w_n)^T$ is called as the blur kernal. From the above homography equation it can be seen that the homography can be determined by knowing the ratio of translation in each of the coordinates and the z value i.e. depth $(t_x/z, t_y/z, t_z/z)$.

Usually in uniform deblurring the values of the rotations and the translation about the z axis are set to zero. This means that we are only considering images that shift in the x or y axis. This is a very constrained problem and can be easily solved.

In non- uniform deblurring it is difficult to not pose any restrictions as the problem would become untractable. This is where different papers pose different restrictions on the variables and define models to solve the different problems. The author has pointed out that there has been till now no other papers have considered out of plane translation i.e a case where t_z is not zero. Thus our method addresses a case where there may be translation about all three axes but the rotation about any axes is assumed to be 0. This assumption is in line with our application where the dash cams or traffic cams will not suffer from any rotational shaking or motion since they will be rigidly secured. On top of this , this paper does not pose any restriction in the direction of the motion vector it does not need to be perpendicular to the sensor plane as assumed in other papers. Thus the motion could be about a slanted direction to the image plane. We achieve this in this paper, by considering the normals of the 3D planes in the image, for example the normal of the billboard, or license plate.

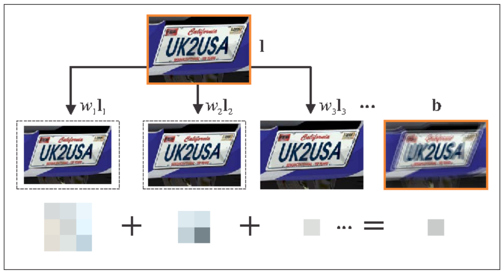

Figure 2: Color mixing due to differences in scale

3 Proposed Approach

3.1 Our Model

The model described in this paper considers any 3D plane in the image to be denoted as $\pi = (n, d)$, where n would be the normal of the plane and d would be the offset of the normal or plane from the camera center, where any point X on the plane would satisfy $X^T \pi = 0$ meaning they are parallel.

In this model images with different camera positions are related by the following homography:

$$H = K(R + tn^T/d)K^{-1} \tag{4}$$

Where d is the depth, since points on a given 3D plane will have varying depth. Since we consider the rotation about the different axes to be 0 , we know that the Rotation matrix will be equal to the identity matrix I. $t = [t_x t_y t_z]^T$ The homography will become:

$$H = K \begin{pmatrix} 1 + t_x n_1 & t_x n_2 & t_x n_3 \\ t_y n_1 & 1 + t_y n_2 & t_y n_3 \\ t_z n_1 & t_z n_2 & 1 t_z n_3 \end{pmatrix} K^{-1} \tag{5}$$

Where $n = (n_1, n_2, n_3)^T$ which is the normal of the 3D plane. Thus given the tranlation values and the normal as input we can determine the homography equation which is used to shift the image from one camera position to the next.

Next the author addresses the problem of Color Mixing represented in Fig(2). When each image is transformed by the homography matrix from the original image to the image at the current position, the scale of each image changes since the image is allowed to move in the z - axis, out of plane translation, this problem is not faced in previous works. But since we are summing the different images at various positions in our current model Eqn(1), this leads to the misrepresentation that every pixel in the blurred image is equal to the sum of a series of corresponding images at different positions. But this is not the case, what happens is that each pixel in the blurred image is actually a sum of different patches of images at various positions, due to the scale one pixel in the blurred image can correspond to a patch of pixels in a sharp image.

This is rectified by changing the model to consider the model of the blurred image to be the sum of sharp images that are blurred by a Gaussian filter. The standard deviation of the Gaussian filter is determined to by the corresponding t_z value. The final blur model is represented as:

$$b = \sum_i w_i P_i^n G_i l + \epsilon \tag{6}$$

Where P_i is the warping matrix that directly corresponds to the homography matrix.

3.2 Implementation

Now that we have our final model for the blurred image, we can define our algorithm for deblurring. From the above equation the known values to us are the blurred image b, the warping matrix P_i,(since we know the normal of the 3D plane, and the translation vector) and the Gaussian kernal G_i. We need to estimate the values of the weight vector w_i and the sharp image l. The author has proposed an alternating minimization to first fix the normal n, estimate the weight w and the image l, then update the value of the normal.

The model is written as:

$$b = \sum_i w_i P_i^n G_i l = B^n l = A^n w \tag{7}$$

where $B^n = \sum_i w_i P_i^n G_i$ and col(i)(A^n) = $P_i^n G_i$

The cost function of the weight vector is given as:

$$E(w) = ||A^n w - b||^2 + \gamma ||w||^2, s.t. \sum_i w_i = 1 \tag{8}$$

Here γ is the smoothing strength. We solve for the value of the weight vector which minimizes the cost function by using Conjugate Gradient method. Since the cost function is a quadratic function it will be a convex function with a minimal value.

Next the value of the sharp image is estimated using sparsity pursuit regularization, where the cost function of the image is given as:

$$E(l) = ||B^n l - b||^2 + \lambda \phi(\nabla l),$$

(9)

Here $\phi(\nabla l)$ is the high sparsity regularization term of the image gradients with L_0 norm. Once the image and weight values are iteratively estimated, the final restored image is produced by optimization by incorporating a natural image hyper-Laplacian.

This algorithm so far uses a fixed normal n which at the beginning was initialized to be the frontal parallel where n = (0,0,-1). In order for the algorithm to take into account the 3D planes in the image we must enable the normal to constantly update. The cost function for the normal is given as:

$$E(n) = || \sum_i w_i P_i^n G_i l - b||^2$$

n = (cos$\alpha sin\beta, sin\alpha sin\beta, cos\beta)^T$(10)

The author has specified that since it is a non linear function, gradient descent will not work. So the normals are sampled by varying the α and β angles of the normal by $15°$ and choose the normal with the lowest cost function.

Thus by constantly updating w and l and update the normal the sharp image can be obtained in 3 iterations in this paper.

4 Experimental Results

Although I clearly understood how the algorithm works, I found a lot of difficulty while implementation. Some of the steps that I took were.

- Created a forward blur example by sampling various translation values i.e. t_z, t_y, t_z, where the z values were sampled more densely. Using these translation values I was able to create a motion blur example for different normal values, where n was (1,0,0),(0,0,1) and (0,1,0). This was achieved by constructing the homography for each image at each position sampled by the t values and adding them together with equal weights to obtain the blurred image. I used functions from the OpenCV library to perform the perspective transform once the homography matrix was calculated.

- Next I constructed the Gaussian matrix G_i for a given standard deviation of [0.1 ,0.5] along the respective axes as mentioned in the paper. The paper did not mention here how the value of the Gaussian matrix should be varied with respect to the t_z value.

- Then I set up the cost equation for the weight kernal update by constructing the A^n matrix and randomly initializing the weights while making sure they added up to 1. I tried to solve this equation for the value of w by using a scipy function scipy.optimize.fmin_cg which is used to minimize a function using non linear conjugate gradient method. I was not completely able to estimate the w values as I had some matrix dimension errors.

- I set up the cost function equation for the image but I was unable to implement a solver for the value of l that minimized the cost function.

5 Conclusion

In this paper , forward motion deblurring was addressed where the motion involves out of plane translation. This solution has applications in several areas such as dash cams and traffic cams. By incorporating the normals of 3D planes in the image, the author has focused on shifting the weights and image values such that the pixels on 3D planes receive higher priority than other pixels in deblurring and they information on the 3D planes will be the most clear. In the paper the author has clearly demonstrated that this approach is novel and has significantly better results for deblurring problems with forward motion. The author also goes onto propose a method to hand label the initial normal values if the 3D planes are too slanted to reduce time of convergence.

Acknowledgments

.

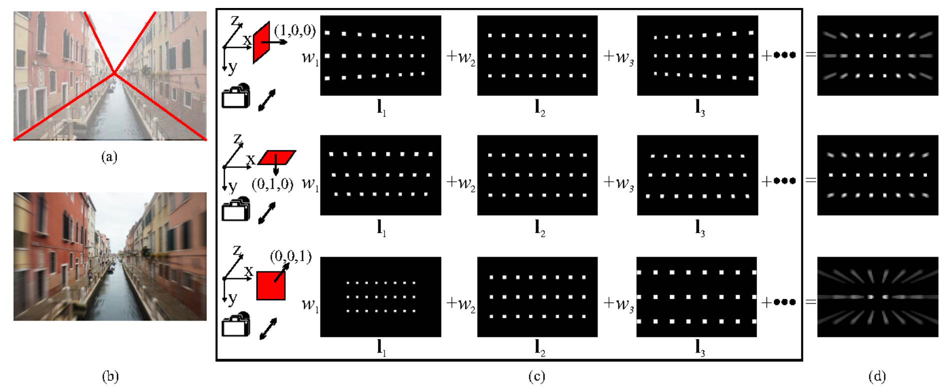

Figure 3. Demonstration of homography bases considering three special normals. (a) Natural images generally contain planar surfaces. (b)

Figure 3: Forward Blur Generation