

# Team 4 - Preliminary Project Report

Tadipatri Uday Kiran Reddy

ee19btech11038@iith.ac.in

Sahukari Chaitanya Varun

ee19btech11040@iith.ac.in

L. Pranay Kumar Reddy

ai19btech11019@iith.ac.in

## Abstract

*With the evolution of high-resolution digital cameras, snapshots have become the most common way to record and share visual data and experiences taken through mobile phones and tablets. But as humans, it is challenging to keep a stiff hold while capturing an image, especially when the image is to be captured on the move; it introduces blur. Though there lot of external tools for frame stabilization, such as gimbal, mounting it on can be challenging at times. Hence, we propose this paper to solve this problem by processing the image with the aid of gyroscope sensed data attached with the camera device, helping to deblur the image. We aim to implement this deblurring with the help of deep convolution nets and extrapolate it with GAN's to give better quality and faster output than traditional CNN's. We can extend this work to stabilise the video feed with lower frame rates with faster computation capabilities.*

## 1. Introduction

Our desired scenes are always in motion; capturing this in the camera would introduce a fuzzy noise known as motion blur. This effect is also very severe in Autonomous vehicles which utilise sophisticated visual SLAM algorithms on live feed coming from cameras or LIDAR sensors that are in motion. Here the motion blur would suppress the crucial features needed to be obtained from the video feed. Even with state-of-the-art technology of camera manufacturing, the motion blur effect is not being improved while the camera resolution is getting better and better.

Mechanically stabilising the sensor like camera or LIDAR is very hard, so we have to use rigorous algorithms to process the image and many recent works like came up with both linear and non-linear models to deblur the image but the processed images are having still some blur and implementation of such sophisticated algorithm in edge devices like mobile phones, DSLR cameras, etc is computationally infeasible. These works leave the fact behind that these motion blurs are caused due to relative motion

between scene and camera. As all the modern mobile phones and camera are equipped with Interrial Measurement Units(IMU), these actually measure velocity, acceleration, we can exploit this sensor data to deblur the images than only using raw image only as data. These **works** actually can up with system to exploit gyroscope sensor data, here they use Convonutaional Neural Network(CNN) with encoder and decoder framework, but these demand high computational power.

In this work we propose a framework which deblurres images with blurred image and gyroscope sensor data, with help of deep-CNN and extrapolate it with Generative Adversarial Network(GAN) to enchanche the quality and faster output of deblurred image.

The below are the main contribution of this paper.

1. We propose a framework known as Motion blur Correction with Gyroscope using Generative Adverserial Networks(MCG-GAN)
2. We use MCG-GAN for deblurring images and show the improvement with existing state-of-art-works.

The rest of the paper is organised as follows - We discuss the detailed problem statement in Section 2. We provide literature review in Section 3 which describes the related works and mentions important results which are going to be used in proposed framework. In section 4, we show results of recent works on deblurring. Finally, in section 5, We cite the references from we have used or described the paper.

## 2. Problem Statement

The problem statement concisely is image deblurring aided by gyroscopic sensing implemented through GAN's, building upon the deep neural net solution. Though the blur introduced can be from relative motion between the camera and picturing object itself, we focus only on deblurring image due to the relative movement of the camera device. The paper DeBlurGAN has already produced a benchmark result in finding the blur and rectifying it, but assumes the blur to be space invariant, i.e, performs blind deblurring over the whole image. We can at first estimate a spatially

variant motion field and blur kernels from a single image using deep networks, thereafter perform non-blind deconvolution. But often, this process requires input of multiple images which might not always be available. We here focus on single image deblurring approach.

From the gyroscopic measurements, the rotation can be estimated as

$$\frac{d\mathbf{q}(\mathbf{t})}{dt} = \frac{1}{2}\mathbf{q}(\mathbf{t}) \odot \omega(\mathbf{t}), \mathbf{q}(\mathbf{t}_1)=1 \quad (1)$$

where  $\omega(\mathbf{t})$  is the 3-dimensional gyroscope measurement and  $\odot$  denotes the quaternion product. The initial condition is given at the starting time of exposure  $t_1$  and the solution is computed at the end time of exposure  $t_2$ . The motion blur caused by translation will depend on the scene depth, which is difficult to estimate from a single image and hence take these limitations into account when generating training data. Consider  $\mathbf{R}(\mathbf{t})$  and  $\mathbf{t}(\mathbf{t})$  as rotation and translation of camera. Assuming the scene depth to be a constant  $\mathbf{d}$ , the motion of camera can be modeled as planar homograph as :

$$\mathbf{H}(\mathbf{t}) = \mathbf{K}[\mathbf{R}(\mathbf{t}) - \frac{\mathbf{t}(\mathbf{t})\mathbf{n}^T}{d}]\mathbf{K}^{-1} \quad (2)$$

where  $\mathbf{K}$  is the intrinsic camera matrix obtained via calibration. By taking the translation to be zero, this simplifies to

$$\mathbf{H}(\mathbf{t}) = \mathbf{K}\mathbf{R}(\mathbf{t})\mathbf{K}^{-1} \quad (3)$$

By this we can compute the blur vector for every pixel, which gives us the blur map of the image  $\mathbf{h}$ .

Now, as we have a blurred image  $I_b$ , our goal is to predict a sharp image  $I$  using GANs, such that

$$I_b = \mathbf{h} * I + \mathbf{n} \quad (4)$$

where  $\mathbf{h}$  is the kernel which models the motion blur also called as blur kernel,  $\mathbf{n}$  is the noise and  $*$  denotes convolution.

### 3. Literature Review

#### 3.1. Motion blur

This is a phenomenon observed in images taken when there is relative motion between camera and the scene, these are the artifacts observed in most common camera which are mostly rolling shutter cameras. Many of the recent works model the motion blur as Linear Space invariant system as follows.

$$I_b = \mathbf{h} * I + \mathbf{n} \quad (5)$$

Where  $I_b$  is Blurred image,  $\mathbf{h}$  is kernel which models the motion blur more oftenly so known as Point Space Function(PSF),  $I$  is actual unblurred image and finally  $\mathbf{n}$  is

noise added while capturing the snap.

In this paper, we use gyroscope data and take into consideration that blur kernel is space variant.

$$I_b = f(I, \mathbf{g}(\mathbf{t}), \mathbf{n}) \quad (6)$$

Where  $f(\cdot)$  is non-linear space invariant function, this is learnt from dataset using GANs which is further more discussed in Section 3.3.

#### 3.2. DeepGyro

The deblurring is based on fully-convolutional neural network which aims to produce a sharp image given blurred image and gyro-based blur field. The network architecture is similar to encoder-decoder network, where the input network consists of blurred RGB image and a gyro-based blur field which pass through series of convolutional and downsampling layers until the lowest resolution is reached. After the bottleneck, this low resolution image is expanded back into a full resolution image with help of upsampling layers. Skip connections are used to allow information sharing between encoder and decoder.

For training of data, the blurred and sharp images with gyro-based blur fields are produced by taking the visual-inertial dataset which consists of various types camera motion resulting in different blur fields. The blur fields are themselves further distorted to exact and noisy blur fields for training a robust model. The paper states of utilizing nearly 100k images and was evaluated on both synthetically and naturally blurred images.

In regard to the current paper we are working on, this referred paper serves as model for extracting the gyro data and also help in generating the datasets required for training and testing the model. This model is embedded in initial stages of our model which is then build further by GANs.

#### 3.3. Generative adversarial networks (GANs)

Generative adversarial networks (GANs), introduced by Ian Goodfellow provides a way to learn deep representations without extensively annotated training data. They achieve this through deriving back-propagation signals through a competitive process also called min-max game involving a pair of networks known as the discriminator and the generator.

The generator receives noise as an input and generates a sample. A discriminator receives a real and generated sample and is trying to distinguish between them. The goal of the generator is to fool the discriminator by generating perceptually convincing samples that can not be distinguished from the real one. The game between the generator  $G$  and

discriminator D is the minimax objective:

$$\min_G \max_D E_{x \sim P_r} [\log(D(x))] + E_{z \sim P_g} [1 - \log(D(G(z)))] \quad (7)$$

where  $P_r$  is the data distribution and  $P_g$  is the model distribution, defined by  $z \sim P_z$ , the input  $z$  is a sample from a simple noise distribution

## 4. Preliminary Results

### 4.1. DeepGyro



Figure 1. DeepGyro performance for increasing levels of blur, Blurred images (green)

## 5. References

1. **Gyroscope-Aided Motion Deblurring with Deep Networks**, Janne Mustaniemi, Juho Kannala, Simo Särkkä, Jiri Matas, Janne Heikkilä
2. **Generative Adversarial Networks**, Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio
3. **DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks**, Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, Jiri Matas