DIGITAL IMAGE PROCESSING

PROJECT REPORT

Synthetic Data for Text Localisation in Natural Images

Joyneel Misra[4] 201401074

Sarthak Sharma[1,2,3] 201431059

Introduction

In this project we introduce a method to generate synthetic images of text in clutter. This engine tried to overlay synthetic text to existing background images in a natural way, trying to accounting local 3-D scene geometry.

Approach

We propose a new method for generating synthetic images of text that naturally blends text in existing natural scenes, using off-the-shelf deep learning and segmentation techniques to align text to the geometry of a background image and respect scene boundaries. We use this method to automatically generate a new synthetic dataset for text in clutter.

Our approach differs entirely from the one proposed by the original authors of the paper. We implement a completely different approach to the given problem statement, building it on the bases of techniques learnt in this course.

Method

Image Segmentation

Segmentation of an image,similar regions being segmented together,was done in by projecting the image to a six dimensional space.

In the higher 6-D space,every point was represented by its unique vector,comprising of (*R,G,B,X,Y,T*) where :

- R, G, B : red, green and blue channel intensity values of the pixel
- X,Y : spatial coordinates of the pixel
- T : texture value in the neighbourhood of the pixel (3rd derivative of the image)

After the projection, the points were clustered according to weighted K-Means algorithm :

$$d_p(y_i, c_k) = \sum_{v=1}^{V} w_{kv}^p |y_{iv} - c_{kv}|^p$$

A higher weight was included for texture and color parameters as the text should respect placed should respect the background it is placed in. A minimal weight was kept for the spatial coordinates so as to retain the spatial locality of the clusters.

Neighbourhood Suppression

Outlier pixels were suppressed by the technique of neighbourhood suppression. For every pixel P_i ,

$$P_i = \begin{cases} P_i & \text{if } \max(\mathbf{W}(P_i))=\min(\mathbf{W}(P_i)) \\ mode(\mathbf{W}(P_i)) & \text{otherwise.} \end{cases}$$

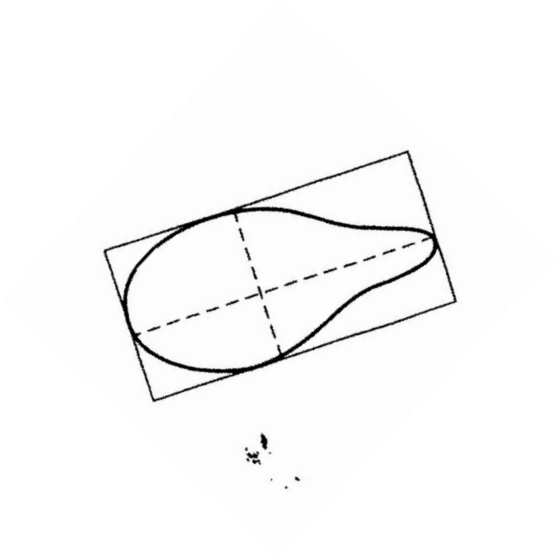
Where $\mathbf{W}(P_i)$ is an array of neighbours of pixel P_i ,including P_i. Neighbourhood of the pixel is characterised by the size of the square surrounding the pixel.

Rectangle Fitting

Naive Way

The naive way is to fit a maximal upright rectangle window in the chosen region for the text to be placed on it. This can be found using a simple Dynamic Programming approach. But the issue that we face is that the text does not respect the orientation of the chosen region and the text image doesn't seem natural.

Improvement 1



We try to find the orientation of the segment by fitting a best-fit rectangular perimeter around the segment. Therefore the orientation will be the slope of the larger side of the rectangle.

So now we rotate the image to make the slope zero, fit the largest rectangle, blend the text image, and un-rotate the image back. But we observe that the issue which is left is that the placed text doesn't respect the depth of the

image, i.e. it doesn't seem that the text is placed on a 3d object.

Improvement 2

We need to consider the depth of each region in the placement of text. Since we do not have access to the depth image, we try to tackle the issue in a supervised setup. We predict the normal of the surface of the region from user input. Then we warp the text according to the surface normal and then blend the warped-text in the image.

Image Blending

For the blending of text with the selected area,Poisson Image Blending was used.Poisson image blending is a technique used for blending two images :

We apply the technique to blend the text with the natural image in the area segmented by the rectangle,to create synthetic data.

Results

Figure 1

Figure 2

Figure 3

References

[1] A. Gupta, A. Vedaldi, A. Zisserman Synthetic Data for Text Localisation in Natural Images IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016

[2] Poisson Image Editing taken from here