Summary of *Image and Depth from a Conventional Camera with a Coded Aperture* by A. Levin, R. Fergus, F. Durand, W.T. Freeman

Raphaël Chekroun

1 Introduction

Conventionals cameras captures blurred versions of scene information away from the plane of focus. This article aims to present a simple modification to existing conventionals cameras to fix this issue.

It consists in the insertion of a patterned occluder within the aperture of the camera lens, thus creating a coded aperture. This allows (i) a high resolution image information and (ii) depth information adequate for semi-automatic extraction of a layered depth representation of the image.

The article starts by introducing a criterion for depth discriminability, used to design the aperture pattern. Then, a stastistical model of images is used to recover both depth information and an all-focus image from single photographs taken with the modified camera. A layered depth map is then extracted, sometimes with the help of the user.

2 Criterion of depth discriminability

First of all, we note that for a planar object at distance D_k the imaging process of cameras can be modeled as convolution:

$$y = f_k * x + n$$

where y is the observed image, x is the true sharp image and f_k is the blur filter, a scaled version version of the aperture shape, and n is noise in neighboring pixels, following a Gaussian model $n \sim N(0, \eta^2 I)$. Thus, the goal here is to determine x.

To do this, we have to use the statistical model of images developped by Olshausen and Field (1996), which take in count the fact that real images have a sparse derivative distribution. We note:

$$P(x) \propto N(0, \Psi)$$

where i, j are the pixels indices, $\Psi^{-1} = \alpha(C_{g_x}^T C_{g_x} + C_{g_y}^T C_{g_y})$, where C_{g_x}, C_{g_y} are the convolution matrices corresponding to the derivative filters $g_x = [1, -1] = g_y^T$. Finally, α is set so the variance of the distribution matches the variance of derivatives in natural images.

We denote $P_k(y)$ the distribution of observed signals under a blur f_k. As the blur f_k linearly transforms the ditribution of sharp images from the previous equation, $P_k(y)$ is also a Gaussian : $P_k(y) \sim N(0, \Sigma_k)$, where $\Sigma_k = C_{f_k} \Psi C_{f_k}^T + \eta^2 I$ is a transformed version of the prior covariance plus noise.

We transform this into the freqency domain, to obtain:

$$P_k(Y) \propto exp(-\tfrac{1}{2} \textstyle\sum_{\nu,\omega} |Y(\nu, \omega)|^2 / \sigma(\nu, \omega)))$$

where $\sigma(\nu, \omega)$ are the diagonal entries of the Fourir transform of Σ_k:

$$\sigma(\nu, \omega) = |F_k(\nu, \omega)|^2 (\alpha(|G_x(\nu, \omega)|^2 + |G_y(\nu, \omega)|^2)^{-1} + \eta^2$$

Intuitively, if the blurry image distribution P_{k_1} and P_{k_2} at depths k_1 and k_2 are similar, it will be hard to tell depths apart. The article then use the Kullback-Leibler (KL) as the criterion of depth discriminability, exprimed here in the frequency domain:

$$D_{KL}(P_{k_1}(y), P_{k_2}(y)) = \textstyle\sum_{\nu,\omega} (\frac{\sigma_{k_1}(\nu,\omega)}{\sigma_{k_2}(\nu,\omega)} - log(\frac{\sigma_{k_1}(\nu,\omega)}{\sigma_{k_2}(\nu,\omega)}))$$

Once the criterion of depth discriminability has been chosen, the shape of the aperture has been taken as the one maximizing it, subject to some conditions: (i) the aperture has to be binary (to make its contruction possible), (ii) it should be feasible from one piece without pieces floating in the middle of it, (iii) do not use the full aperture in order to reduce radial distortion, and (iv) the sizes of the hole in the filter have to be bigger than a limit size to avoid diffraction.

3 Recover an image captured with such an aperture

3.1 Deblurring the image

Once the correct blur scale of an image y is identified, we want to remove the blur to reconstruct the sharp, original image x. This step of *debluring* can be noted, under our probalistic model:

$$P_k(x|y) \propto exp(-\tfrac{1}{\eta^2}|C_{f_k}x - y|^2 + \alpha(|C_{g_x}x|^2 + |C_{g_y}x|^2))$$

Thus, we're searching here:

$$x^* = argmin(\tfrac{1}{\eta^2}|C_{f_k}x - y|^2 + \alpha(|C_{g_x}x|^2 + |C_{g_y}x|^2))$$

which is the x minimizing the reconstruction error $|C_{f_k}x - y|^2$, with the prior preferring x to be as smooth as possible. This can be done by solving a set of linear equations, not detailed here.

3.2 Handling depth variations

All this work was done supposing the object is planar and at a constant distance of the camera, so the blur kernel would be uniform on the image. The reality is different: a true image contains informations to keep on different depths.

To take account of this, the article use small local windows within which the depth is assumed to be constant. Nonetheless this highlight another issue: as the windows are small, the depth classification might be unreliable.

To take in count depth variation, the first thing to do is to compute K decoded images, $x_1 ... x_K$. Then, compute the reconstruction error $e_k = y - f_k * x_k$. A decoded image x_k will provide smooth possible reconstructions for parts of the image where k is the true scale. In other regions, some artefacts may appear, and their presence ensure that the reconstruction error will be high.

Then, we compute on every pixel i an approximation of the local energy around this pixel, by averaging the construction error over a small local window:

$$E_k(y(i)) = \sum_{j \in W_i} e_k(j)^2$$

This local estimate is finally used to locally select the depth $d(i)$ of the i-th pixel:

$$d(i) = argmin_k \lambda_k E_k(y(i))$$

To obtain a good image, the deblurred image is constructed such that, for every pixel i, $x(i) = x_{d(i)}(i)$.

Nonetheless, to produce a depth estimate which could be useful for tasks like object extraction and scene re-rendering, the depth map has to be smoothed. To do this, the article searches a close to d but smoothed d^* with an energy minimization method using a Markov random field over the image.

4 Results and conclusion

The resultats are pretty convincing, as we'll see during the presentation. Nonetheless, it occasionally happens than the depth labeling misses the exact layer boudaries, due to insufficient contrast. To correct this, the user may apply a brush strokes to the image with the required depth assignement.

This paper has evident direct applications: it allows to, without adding expensive and complex chips or objectives, take all-focus pictures and build depth-map of pictures.

This system is nonetheless still not perfect: the insertion of the lens reduces by 50% the amount of light that reaches the sensor, and the blur filter must be perfectly calibrated over depth value.