Depth reconstruction by means of defocused light from incomplete measurements

A. Zakharova[1], O. Laligant[2] et C. Stolz[2]

[1] INSA Rouen, LMI, Avenue de l'Université - 76801 Saint-Étienne-du-Rouvray, France
anastasia.zakharova@insa-rouen.fr

[2] Le2i Lab., Université de Bourgogne, 12 rue de la Fonderie, 71200 Le Creusot

Abstract

We present a novel method of depth estimation based on the defocused light. The illuminated pattern is projected onto the object and we estimate the diameters of blur spots on the object (assuming that the depth linearly depends on the diameter of the spot). To reconstruct the depth map of the object, we suppose that its gradient is sparse — in this case, we show that one can reconstruct the depth map from incomplete measurements. The idea comes from the compressed sensing theory : however, since the measurements are not linear it cannot be directly applied in our setup. Thus we develop a methodology of reconstruction from incomplete sparse nonlinear measurements and apply it to $3d$ *reconstruction from blurred images.*

Key words

Depth from defocus, sparsity, ℓ_1-minimization

1 Introduction

The aim of our research is the 3-dimensional reconstruction of the object (or calculating its depth map) from given $2d$ measurements.

To estimate the depth the technique based on the defocused light is used. That is, we project a given pattern on the object ; then, if the image of the pattern is not in the focal plane, we get a blurred image of the pattern on the object. The blur level at each point depends on the distance from the focal plane ('depth of the point'), thus we can estimate the depth by measuring the blur around the point. Since the depth is not defined uniquely by the blur level (the blurred point could be either before or behind the focal plane), one needs at least 2 measurements obtained by projecting a pattern from different distances in order to define the depth. Having a series of such blurred images obtained by projecting a selection of patterns from different distances one can reconstruct the depth map of the object.

There exist several methods of depth computation, see [1][2][3]. Extension of these methods take benefit of active lighting to improve accuracy and avoid some limitations [4][5][6]. However, they are all based on the analysis of the acquired blurred $2d$ images which could be of large size, thus depth estimation appears to be quite costly. That is why we want to benefit from the compressed sensing ideas. This domain is highly developed in the past several years both in theoretical and application sense. It claims that if we have an object (a signal or an image) which is either sparse in some basis (that is, there are few non-zero coefficients in its representation in this basis) or compressible (could be well approximated by a sparse one) we can exactly reconstruct the object from a number of global measurements less than the dimension of the object itself, i.e. from incomplete measurements, by solving a minimization problem. We propose to change the acquisition process in order not to measure all the blurred image but take a linear combination of the values in some points instead. This means that each projected pattern results in one scalar measurement only.

There are two main difficulties arising in such a framework. Firstly, we are not able to take direct linear measurements of the depths from a blurred image of the pattern since we do not know their value. To compute it, we first estimate the intensity in this point and then compute the depth as its inverse (up to a multiplication coefficient). That is, we measure linear combinations of intensities then reconstruct the intensity map and compute the dense map. Since we suppose that depth is sparse but we apply the reconstruction algorithm to the intensity then the sparsity basis must be chosen in such a way that intensity was also sparse. By choosing the total variation minimization we will satisfy this constraint.

Secondly, the matrices which are commonly used for measurements are dense since in order to compensate the lack of information due to the limited number of measurements one prefers to measure all the object during each measurement. In the contrary, in our framework the projected pattern must be sparse enough, since otherwise the circles of confusion resulting from object points with different depths will intersect and this will considerably complicate the evaluation of the depth in a given point. Thus we are able to make a depth estimation in a limited number of points only. A possible solution is to reconstruct from the mea-

surements obtained with sparse measurement matrices but this will require a considerable amount of measurements. Thus we will reconstruct the $3d$ objects in lower resolution than one of the given $2d$ images. This reduces significantly the number of measurements and we will see that it gives appropriate results.
In the following we explain how we need to adapt the ideas of compressed sensing to our case. The outline is as follows : we start by setting out the basic compressed sensing ideas and explain why they cannot be applied directly to the acquisition process. Nevertheless, we show that under some modifications of the classical compressed sensing theory we can reconstruct the depth map of the object using incomplete $2d$ measurements. Then we describe the method of depth estimation by means of defocused light together with the experimental framework : the optical system design and its limitations. To conclude, we present the reconstruction results both for synthesized objects and for the real objects using the measurements acquired by the optical system based on defocused light.

2 Sparse recovery

The idea of sparse recovery consists in reconstruction of a signal from a number of *non-adaptive* measurements that is much lower than the dimension of the signal. The conditions imposed on the signal are that the signal is sparse, i.e. the information about it is contained in the very few signal components ; in the contrast, each measurements should contain information about all the signal components ; in other words, the signal must be *localle distributed* and measured with *global* measurements.
The theory below concerns the reconstruction in dimension one ; in case of a two-dimensional image we write it as a one-dimensional vector by concatenating its columns.
Let x be a vector in $\mathbb{R}^n$. We say that x is *s-sparse* if it has at most s nonzero components. In other words, if we define

$$\|x\|_0 := |\text{supp}(x)|,\ \text{supp}(x) = \{j\ :\ x_j \neq 0\}$$

then the vector is s-sparse if $\|x\|_0 \leq s$. However, in practice we usually deal with compressible vectors, i.e. vectors which are close to sparse ones in the sense of a norm.
In order to reconstruct we take $m \ll n$ linear measurements of x :

$$y_i = a_{i1}x_1 + \ldots + a_{in}x_n, i = 1, \ldots, m,$$

which is equivalent to applying to x a matrix $A \in \mathbb{R}^{m \times n}$, called the measurement matrix (each line of A corresponds to one measurement) and results in a measurement vector $y \in \mathbb{R}^m$:

$$y = Ax.$$

To recover x from incomplete measurements y, we are looking for the sparsest vector which is consistent to measurements y ; a natural way to do it is to solve the following minimization problem (also called ℓ_0-minimization problem)

$$\min \|x\|_0 \text{ subject to } Ax = y. \tag{1}$$

However, since this problem is not convex and it is NP-hard in general, it cannot be used in practice. It is thus replaced by either ℓ_1-minimization problem (or basis pursuit)

$$\min \|x\|_1 \text{ subject to } Ax = y, \tag{2}$$

which is a convex optimization problem, or by greedy algorithms (matching pursuit etc.). Under the appropriate conditions on the matrix A and the sparsity of x the solution obtained by these approaches coincides with the solution of (1).
In this paper we will concentrate on the reconstruction by ℓ_1-minimization. The most widely used sufficient condition on the matrix A is the so-called *restricted isometry property* (RIP). One says that a matrix A satisfies the RIP of order s (later RIP-s) if there exists $\delta_s \in (0, 1)$ such that for all s-sparse vectors z it holds

$$(1 - \delta_s)\|z\|_2^2 \leq \|Az\|_2^2 \leq (1 + \delta_s)\|z\|_2^2.$$

It is well known ([7]) that if a matrix satisfies RIP-s, then all the s-sparse vectors will be recovered and stability in l_1 also takes place. In this case, one needs $Cs \log \frac{n}{m}$ measurements for en exact reconstruction (see for example [8]). Most of the known examples of RIP matrices are given by random — Gaussian or Bernoulli — matrices. There exist however also deterministic matrices satisfying RIP ; as examples, we note partial Fourier, circulant and Toeplitz matrices.
Let us now pass on to the depth reconstruction. Let X be an (unknown) matrix of dimension $N = n \times n$, A be the measurement matrix, E the additive noise and Y the resulting measurement vector. We adopt a recovery model which assumes that the gradient of this function is sparse. It means that we minimize the total variation which is the sum of the magnitudes of the gradient at every point :

$$TV(x) := \sum_{i,j} \sqrt{(D_{ij}^h)^2 + (D_{ij}^v)^2}$$

where

$$D_{ij}^h = \begin{cases} X_{i+1,j} - X_{ij},\ i < n \\ 0\ i = n. \end{cases}$$

and

$$D_{ij}^v = \begin{cases} X_{i,j+1} - X_{ij},\ j < n \\ 0\ j = n. \end{cases}$$

Then we solve the following minimization problem

$$\min \|TV\|_1 \text{ subject to } \|AX - Y\|_2 \leq \varepsilon$$

where ε corresponds to the standard deviation of noise E.
The choice of the TV-minimization model is very common for image reconstruction but it is also governed by the specific of our problem. As it was already mentioned, we cannot directly apply this model to the optical system since we do not dispose depth values ; we compute the intensity map whereas the sparsity assumption is made with respect to the depth map. The advantage of the TV-sparsity is that when the depth gradient is sparse, the intensity gradient will also be sparse thus we can apply the described setting.

3 Optical system design

Figure 1 presents the principle of the current version of the optical system and an example of image acquired by the sensor. The path of the light is as follows :

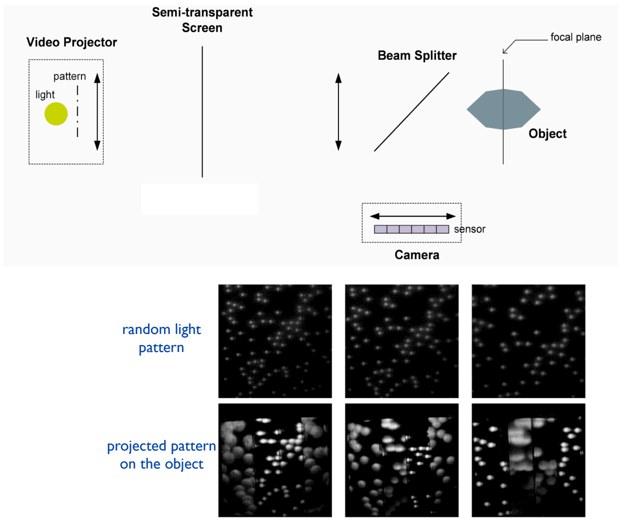

Figure 1 – *Principle of the optical system.*

- the video projector generates a dynamic random light pattern
- the pattern is projected on a semi-transparent screen to obtain a magnified pattern
- the magnified pattern is then projected by a lens on the 3D object
- the projected pattern being reduced in size, the depth of field is tight and induces sharp and blurred spots on the object
- the pattern on the object is observed with the image sensor thanks to the beam splitter

The pixels are selected according to the generated pattern : if the projected point is sharp at a given location, the pixel corresponds to a maximum of light. If the projected point is blurred, the pixel corresponds to a point on the spot profile (the center of the spot is delocalized). We finally obtain a measure by summing the luminance of the selected pixels. In this way, we simulate a one-pixel camera with a selecting grid controlled by the dynamic pattern.
Let us now discuss the appropriate patterns. The aim is to avoid the patterns that are too dense since otherwise the circles of confusion will overlap and the depth estimation would no more be possible. We propose then to use dot patterns with dots being sufficiently far one from another. However, with such a pattern we obtain a depth estimation for a limited number of points only. We will see later that these measurements still allow us to reconstruct all the depth map.

4 Experimental results

Let us denote by $X = (X_{ij})_{i,j=1}^{n}$ the depth map. We will later assume that X is written as a vector of length $N = n^2$. In order to simplify we consider here a non-ambiguous case when all the object is behind the focal plane, i.e. the correspondence intensity-depth is one-to-one.
For simulation, let us consider a 6-step stairs ; the resolution of taken images is 128×128. The modeled projected pattern consists of $M = 20$ points (M must be much smaller than the dimension of the object). The locations of the pattern light points are chosen (pseudo)randomly in such a way that the blur circles do not overlap. Then in each point of the image we model the intensity of the image I by a Gaussian

$$\frac{1}{\sqrt{2\sigma_i^2}} exp\left(-\frac{(t-c_i)^2}{2\sigma_i^2}\right),\ i = 1, \dots, M$$

with the maxima in the points c_i that correspond to the light points of the pattern and with the variances σ_i corresponding to the distance between the object point and the lens. We suppose that the correspondence depth-blur is given by previous calibration and that the locations of maxima are known (since for each measurement we know the projected pattern). Then we're interested only in the intensity maxima.

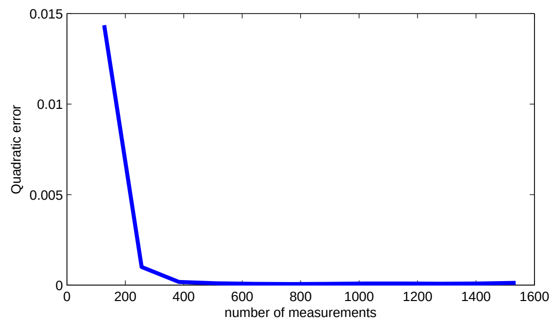

Figure 2 – *Dependance of the error from the number of the measurements ; we see that roughly 400 measurements suffice to reconstruct a depth map of dimension* 128×128.

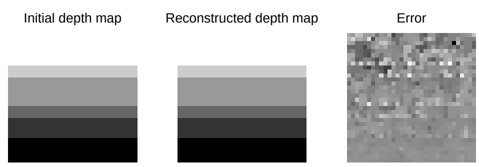

Figure 3 – *Reconstruction results from 0.02% of the data ; different gray levels present different depths.*

As shown on the Figure 2 the l2-error between the depth map of the initial object and the reconstructed depth map

is a decreasing function of the number of the measurements which stabilizes from roughly 400 measurements (i.e. 0.02% of measured data). The reconstruction result is presented on Figure 3 : one can see that for a human eye the initial and the reconstructed depth maps look exactly the same.
Similar tests were also performed with respect to a ball ; due to the fact that the total variation of the stairs object is sparser than one of a ball the reconstruction of the depth map of this latter needs more measurements ; still, it appears to be possible to reconstruct it from incomplete measurements.

5 Conclusion

In this paper we briefly presented the idea of reconstruction from incomplete measurements and explained how it can be applied to the problem of 3d reconstruction by means of defocused light. The simulated examples justify our approach. The work on the real images is in progress.

6 Acknowledgements

The authors acknowledge the support of the European Union, projet no 33760, and of the Burgundy Region.

Références

[1] J. Ens et P. Lawrence. An investigation of methods for determining depth from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(2) :97–108, 1993.

[2] S.K. Nayar et Y. Nakagawa. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8) :824–831, 1994.

[3] Y. Xiong et S.A. Shafer. Depth from focusing and defocusing. Dans *IEEE Conference on Computer Vision and Pattern Recognition*, pages 68–68. Citeseer, 1993.

[4] M. Noguchi et SK Nayar. Microscopic shape from focus using a projected illumination pattern. *Mathematical and computer modelling*, 24(5-6) :31–48, 1996.

[5] I. Lertrusdachakul, Y.D. Fougerolle, et O. Laligant. A novel 3D reconstruction approach by dynamic (de) focused light. Dans *Proceedings of SPIE*, volume 7538, page 75380L, 2010.

[6] Intuon Lertrusdachakul, Yohan D. Fougerolle, et Olivier Laligant. Dynamic (de)focused projection for three-dimensional reconstruction. *Optical Engineering*, 50(11) :113201–113201–11, 2011.

[7] T. Tao E. Candès et J. Romberg. Robust uncertainty principles : exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2) :489–509, Mars 2006.

[8] E.J. Candès. The restricted isometry property and its implications for compressed sensing. *Sci. Paris*, 346(9-10) :589–592, Mars 2008.