Environment Matting by Sparse Recovery

Bálint Fodor

Introduction

Environment matting is an image based relighting technique that aims to extract information about the light transport of a given scene exploiting the assumptions of the measurement setting. The measurement involves a digital camera and a controllable screen as the background of the scene.

While displaying special patterns in the background the camera captures the light response of the scene. Processing such measurements a background-free representation of the scene is aimed to be produced. The technique of environment compositing is to offline *re*render the scene with arbitrary background using the extracted representation.

Based on recent results in the field of compressive sensing and sparse signal recovery [1] this work solves the problem of environment matting by assuming that only few background pixels are linearly interplaying to produce a single camera pixel.

Thus a vector of corresponding pixel intensities can be thought of as a linear combination of some entries of a known signal dictionary built from vectors of corresponding pixel intensities through the displayed background patterns.

Previous Work

Whilst related works on image matting deal with basically opaque scene elements [4] the concept of environment matting intruduced by Zongker *et al.* [7] is intended to reveal the reflective and refractive nature of the scene.

Wexler *et al.* [5] formulates and solves a more general problem where the background varies due to camera movement. It involves two steps: first a merged pure background is extracted, then the 2D-2D map of the light transport is estimated. Experiments show that the method has limited accuracy especially when employing non-planar backgrounds.

Frequency and wavelet based approaches [6] are more robust but require a large number of images making the measurement last long.

Efforts on improving accuracy and time-consumption are also made [2]. In [2] the mapping between the background and the camera image is estimated by color-coding all the pixels in the backgound image and matching those of in the camera image.

The most related method of this work is attributed to Duan *et al.* [3]. A compressive sensing scheme is used to estimate the contribution coefficients of pixels by displaying patterns in a coarse-to-fine manner. It is assumed and exploited that the background pixels supporting a camera pixel (i.e. pixels that have non-zero coefficient) are spatially grouped in the background image.

Methodology

It is assumed that there is a linear relationship between the pixels of the screen and that of the camera. So the common gamma function-like intensity distortion of the devices are compensated.

The number of camera and screen pixels are denoted by d_c and d_s respectively.

A pixel value of the camera is modelled by a linear combination of at most k screen pixels:

$$c_i = \sum_{j=1}^{d_s} \alpha_j s_j, \quad |\{j : \alpha_j \neq 0\}| \leq k \tag{1}$$

where c_i is the ith pixel of the camera, s_j is the jth pixel of the screen and $0 \leq \alpha_j \leq 1$ are the screen pixel coefficients. The aim is to extract the k non-zero α_j coefficients.

Displaying M patterns (1) can be easily rewritten as $C = AS$ where C is $d_c \times M$, S is $d_s \times M$ matrices and A is a $d_c \times d_s$ sparse matrix. Each corresponding coloumns of S and C represent a pattern-response image pair. As A is the coefficient matrix it has at most k non-zero entries in each row.

Compressed sensing theory suggests employing random patterns for the problem of sparse recovery so the entries of S are choosen to be realizations of a random variable with normal distribution. Here S is called the signal dictionary. For every camera pixel a row of C belongs.

Given x a row of C the problem is to find k rows (k non-zero coefficient) in S that all together are the best approximation of x.

A is reconstructed via the Orthogonal Matching Pursuit algorithm. It is a relatively simple, greedy method and iteratively decreases the residual error by selecting the vector from the dictionary that is maximizing the dot product with the residual. The coefficients of the so far selected entries are updated in each step.

Finding the first k entries and their coefficients gives the sparse reconstruction of x. The pixels corresponding to the selected entries of S are treated as the main contributors that form x.

In this work the problem of environment matting is considerd employing a compressive sensing approach. Using OMP as the sparse solver the performance of the concept is evaluated and its competitiveness is justified.

References

[1] E.J. Candès and M.B. Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, 2008.

[2] B. Choudhury, D. Singla, and S. Chandran. Fast color-space decomposition based environment matting. In *Proceedings of the 2008 symposium on Interactive 3D graphics and games*, page 1. ACM, 2008.

[3] Qi Duan, Jianfei Cai, Jianmin Zheng, and Weisi Lin. Fast environment matting extraction using compressive sensing. In *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pages 1 –6, july 2011.

[4] J. Wang and M.F. Cohen. Image and video matting: a survey. *Foundations and Trends® in Computer Graphics and Vision*, 3(2):97–175, 2007.

[5] Y. Wexler, A.W. Fitzgibbon, and A. Zisserman. Image-based environment matting. In *Proceedings of the 13th Eurographics workshop on Rendering*, pages 279–290. Eurographics Association, 2002.

[6] J. Zhu and Y.H. Yang. Frequency-based environment matting. In *Computer Graphics and Applications, 2004. PG 2004. Proceedings. 12th Pacific Conference on*, pages 402–410. IEEE, 2004.

[7] D.E. Zongker, D.M. Werner, B. Curless, and D.H. Salesin. Environment matting and compositing. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 205–214. ACM Press/Addison-Wesley Publishing Co., 1999.