An SNR analysis of structured light techniques in the presence of strong ambient illumination

Supplementary Technical Report
for Paper ID 0100

In this technical report, we provide an analysis on the image formulation of structured light based techniques in the presence of strong ambient illumination. In Section 1, we derive an analytical expression for the signal-noise-ratio (SNR) of structured light coding schemes. We show that the SNR is proportional to the power ratio between structured light signal and ambient illumination. In Section 2, we formulate the theoretical relation between depth accuracy and SNR. Based on this relation, the concept of decodability condition is present for general coding schemes. In Section 3, we derive a theoretical lower bound on acquisition time for different structured light coding schemes.

1 SNR formulation of structured light coding scheme in strong ambient illumination

Structured light methods belong to the category of active triangulation techniques. Its general setup consists of a strucutred light source (projector) and a camera. We model the structured light source L as a projector that has an image plane with C columns. The projector projects spatio-temporally coded patterns on the scene so that a unique intensity code is assigned to each column [1].

The intensity of a scene point S in a captured image is:

$$I = I_l + I_a + \eta, \tag{1}$$

where I_l and I_a are intensities corresponding to the light source L and ambient illumination A, respectively. η is the camera noise. The goal is to extract the signal component I_l reliably from the captured images.

The accuracy of the estimated signal $\hat{I}_l$ is proportional to the signal-to-noise-ratio:

---

[1]Because of epipolar geometry between the projector and camera, only 1D coding (e.g., along the columns) on the projector plane is sufficient to perform depth recovery using triangulation.

$$SNR = \frac{\Delta(I_l)}{\eta} = \frac{1}{L-1} \cdot \frac{I_l}{\eta}, \tag{2}$$

where $\Delta(I_l)$ is the minimal distance required to distinguish two different signals. L is number of different itensities used for coding. For example, since binary methods [3] only use two different intenties, $L = 2$ in this case. For N-ary coding [1] or color coding [4], L equals to N or the color number.

The components I_l and I_a are proportional to the irradiance values R_l and R_a at scene point S due to the light source L and ambient illumination A, respectively:

$$I_l = \alpha R_l, \quad I_a = \beta R_a, \tag{3}$$

where α and β encapsulate the scene point's BRDF, light fall-off, and camera gain. β also includes the effect of any optical (e.g., spectral or polarization) filtering used for reducing ambient illumination.

We assume the affine camera noise model, with both signal-dependent and signal-independent terms [2]:

$$\eta^2 = \sigma_r^2 + \frac{\alpha R_l + \beta R_a}{g}, \tag{4}$$

where σ_r is the standard deviation of the signal-independent sensor read noise, and g is camera gain. In scenarios with strong ambient illumination, $R_a >> R_l$, and the dominant source of noise is the signal-dependent photon noise, i.e., $\sigma_r^2 << \frac{\beta R_a}{g}$. Then, the SNR is approximated as:

$$SNR \approx \lambda \cdot \frac{1}{L-1} \cdot \frac{R_l}{\sqrt{R_a}}, \tag{5}$$

where $\lambda = \sqrt{\frac{\alpha^2 g}{\beta}}$. The reflection properties of the scene are assumed to be the same for structured light and ambient illumination. Also, we used a narrow-band laser light source and spectral filter in front of our camera. This suppresses ambient illumination by a factor of 20. In specific, $\alpha = 0.250$, $\beta = 0.0125$, $g = 4.0$ for our setup and λ is caculated as 4.47 accordingly.

2 Theoretical relation between SNR and depth accuracy

In this section, we derive the expression of depth accuracy given a SNR level. We have known that this relation depends on the structured light coding and decoding algorithms. For analysis simplicity, we assume that minimum-distance decoding is used for all coding schemes. Also, we assume the smallest hamming

distance of a coding scheme is always 1, which means the signal can be decoded correctly only when all the code digits are completely correct.

Suppose I_l^i is an intensity code, the captured intensity is $\hat{I}_l^i = I_l^i + \eta$ due to sensor noise. The condition that I_l^i can be decoded correctly with minimum-distance decoder is:

$$|\hat{I}_l^i - I_l^i| < |\hat{I}_l^i - I_l^j|, \text{ for all } j \neq i \tag{6}$$

For Eq. 6 to hold for even the nearest two codes, we get the condition:

$$\min(\frac{I_l^i - I_l^j}{2}) > \frac{\tau}{2} \cdot \eta, \text{ for all } j \neq i \tag{7}$$

where τ is a parameter to control the probability of decoding correctness. Since photon noise is the dominating noise source in strong ambient illumination, we use a gaussian distribution $N(\cdot)$ to approximate the noise distribution of each intensity code. Therefore, we can give the probability $p_r(i)$ that a single intensity code is decoded correctly:

$$p_r(i) = N(\min(\frac{I_l^i - I_l^j}{2}) > \frac{\tau}{2} \cdot \eta) = N(\frac{1}{L-1} \cdot I_l > \tau \cdot \eta) = N(\text{SNR} > \tau) \tag{8}$$

where L is the number of intensity levels as discussed in Section 1.

We assume the code length of a specific coding scheme is N_c and all code bits are independent to each other. The probability that the code sequence is correctly decoded is:

$$P_r = \prod p_r(i), \, 1 < i < N_c \tag{9}$$

The error tolerance of correspondence decoding is used as the measurement of depth accuracy.

$$\delta = \mu \cdot (1 - \prod p_r(i)), \, 1 < i < N_c \tag{10}$$

where μ is a weight constant set to be 10 in our setting.

In order to achieve a desired depth accuracy δ, the SNR should be higher than a threshold τ, i.e., $SNR > \tau$. The threshold τ increases monotonically with δ. Substituting in Eq. 5:

$$\frac{1}{L-1} \cdot \frac{R_l}{\sqrt{R_a}} \geq \frac{\tau}{\lambda}. \tag{11}$$

We call this the *decodability condition*. In order to achieve the desired depth accuracy, all the captured images must satisfy the decodability condition. In our experiment, we set the decoding error tolerance to be 0.5 column in average. As a result, the τ equals to 3.0 in our experiment.

3 Theoretical lower bound on acquisition time of structured light coding schemes

In this section, we discuss the theoretical lower bound on the aquisition time of existing structured light coding schemes. As introduced in Section 2, to achieve a required accuracy level, structured light coding schemes have to meet the decodability condition.

When the SNR is insufficient to hold the decodability condition, a common technique for increasing SNR is by capturing multiple frames per image. It will increase the SNR by a factor of $\sqrt{f}$ when f frames are averaged per image. To meet the decodability condition using frame-averaging, we get the expression for the minimal number of frames per image:

$$f \geq \left(\frac{\tau \cdot (L-1)}{\lambda R_l} \right)^2 R_a \ . \tag{12}$$

Let N_c be the number of images required by the particular structured light coding scheme used to encode all the projector columns uniquely, and f, as defined above, is the number of frames to be averaged per image. Then, the total number of measurements M is given as:

$$M = N_c \times f \ . \tag{13}$$

If we set a constant exposure time to all the captured frames, we will get a lower bound on the total number of measurements, which is also a lower bound on the aquisition time:

$$M \geq N_c \cdot \max(\left(\frac{\tau \cdot (L-1)}{\lambda R_l} \right)^2 R_a, 1) \ . \tag{14}$$

Eq. 14 is applicable to general existing structured light schemes. For example, we set the $L = 2$, then we get the lower bound for binary coding. We set the L to be the number of intensity levels in N-ary or color coding, then we also get the lower bound for them.

In our paper, we introduce the concept of light redistribution. Then we bring in a new dimensoin in the Eq. 14. If the total column number is C, and K columns are illuminated at a time, Eq. 14 can be re-written as:

$$M \geq \frac{C}{K} \cdot N_c^{'} \cdot \max(\left(\frac{K \cdot \tau \cdot (L-1)}{C \cdot \lambda R_l} \right)^2 R_a, 1) \ . \tag{15}$$

For more discussion on the light redistribution, please check the paper for more details.

References

[1] D. Caspi, N. Kiryati, and J. Shamir. Range imaging with adaptive color structured light. *IEEE PAMI*, 20(5), 1998.

[2] Samuel W. Hasinoff, Fredo Durand, and William T. Freeman. Noise-optimal capture for high dynamic range photography. In *CVPR*, 2010.

[3] J. Posdamer and M. Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18(1), 1982.

[4] L. Zhang, B. Curless, and S.M. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. *Proc. IEEE 3DPVT*, 2002.