ON THE ACCURACY OF POINT LOCALISATION IN A CIRCULAR CAMERA-ARRAY

Alireza Ghasemi   *Adam Scholefield*   *Martin Vetterli*

School of Computer and Communication Sciences
École Polytechnique Fédérale de Lausanne

ABSTRACT

Although many advances have been made in light-field and camera-array image processing, there is still a lack of thorough analysis of the localisation accuracy of different multi-camera systems. By considering the problem from a frame-quantisation perspective, we are able to quantify the point localisation error of circular camera configurations. Specifically, we obtain closed form expressions bounding the localisation error in terms of the parameters describing the acquisition setup.

These theoretical results are independent of the localisation algorithm and thus provide fundamental limits on performance. Furthermore, the new frame-quantisation perspective is general enough to be extended to more complex camera configurations.

Index Terms— Image sampling, Image sequences, Reconstruction algorithms, Signal quantization.

1. INTRODUCTION

A key advantage of multi-camera systems is their ability to recover the 3-D information lost during projection. However, a formal, quantitive analysis on the accuracy of such recovered information is still lacking. This is surprising given the long-standing interest in these systems [1, 2, 3] and improvements that they can make to vision algorithms [4, 5, 6, 7, 8].

Recently, a range of novel camera-array architectures, such as the Lytro light-field camera [9] and PiCam mobile camera array architecture [10] have been introduced, altering the direction of modern consumer digital photography. In the presence of such developments a rigorous treatment of this topic is needed.

In the two-view (stereo-imaging) case, probabilistic error analyses have been carried out [11, 12]; however in the case of more than two views, no such analysis exists. Excellent multi-view scene reconstruction algorithms have been proposed [13, 14], but they have not considered a quantitative error analysis. Chai et al [15] studied the frequency spectrum of the plenoptic function [16], which has been used to find the critical sampling rate for error-free image-based rendering of different classes of scenes [17, 18, 19]. Raynor et. al. [20] analysed the error in range finding using a plenoptic camera utilising a model based on Gaussian noise.
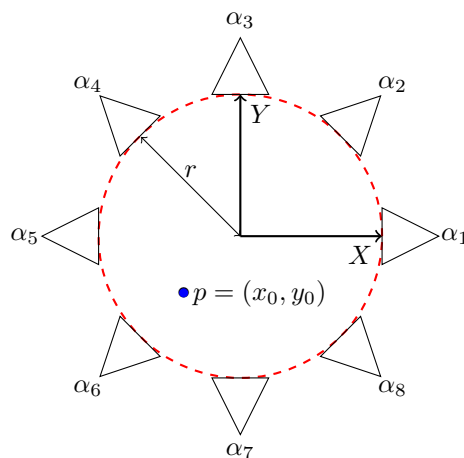
Fig. 1: A point $p = (x_0, y_0)$ in the unit-circle is captured by $m = 8$ cameras uniformly spaced around a circle. The problem is to localise this point from these m projections.

In this paper, we study circular camera-arrays, in an algorithm-independent way, through different error criteria. We relate the problem to frame quantisation [21, 22, 23, 24], which is well studied in the information theory community and has a rigorous mathematical foundation. In particular, Cvetković [25] has analysed the partitions induced by uniform scalar quantisation of expansions of $\mathbb{R}^2$.

We show that these results can be utilised to bound the localisation error of circular camera systems under orthogonal projections. Furthermore, we extend these results to the perspective projection case.

More precisely, we show that, when the number of cameras is sufficiently large, the localisation error is upper bounded by a term that decreases quadratically with respect to the number of cameras. This shows that the error tends to zero as the number of cameras tends to infinity. These results hold for the majority of the region of interest, however there is an exception, which we fully quantify. Finally, we provide numerical simulations, which verify the quadratic bound.

2. PROBLEM SETUP

We are interested in the error in reconstructing a 3-D point from different multi-camera configurations. For simplicity, we will limit our analysis to circular-camera setups, such as the configuration shown in Fig. 1. Furthermore, we will as-
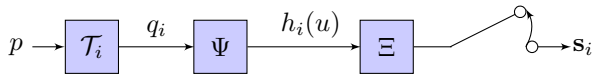
Fig. 2: Camera acquisition model.

sume that the world is 2-D, rather than 3-D, resulting in 1-D images. This allows us to easily visualise the results and the extension to 3-D is relatively straight forward.

More formally, we wish to reconstruct a point $p = (x_0, y_0)$, located within a circle of radius r, from m images, captured with m cameras positioned uniformly on the perimeter of the circle and oriented towards the origin. We will refer to the interior of this circle as the region of interest. Let us index the cameras from 1 to m and associate each one with its angular location α_i, measured anti-clockwise from the X-axis. Figure 1 depicts this configuration for the case of eight cameras.

Our goal is to find an estimator, $\hat{p} = (\hat{x_0}, \hat{y_0})$, for the location, $p = (x_0, y_0)$, and analyse its behaviour and error bounds. To do this we need to understand how the point p is captured in each camera. To this end, we utilise the camera acquisition model depicted in Fig. 2.

Here, $\mathcal{T}_i$ is a projection operator, which maps the 2-D point p onto the 1-D image plane of the i-th sensor. The 1-D projection is then subjected to a convolution with a point-spread function (PSF), Ψ, which models the blurring of the camera lens. Finally, the continuous-domain to discrete-domain sampling is achieved by a convolution with the integration kernel, Ξ, before ideal sampling. The details and justification of each of these steps is given in the following subsections.

2.1. The Projection Operator

The first and most critical stage of the acquisition model is the projection of the point p onto the image plane of the i-th sensor. This image plane is the line tangent to the circle at the point $[\cos \alpha_i, \sin \alpha_i]^T$.

We will consider both orthogonal and non-orthogonal projections, as depicted in Fig. 3.

The orthogonal projection case has been widely studied and lays the foundations of computed tomography and magnetic resonance imaging. In this case, the projection point, q_i^o, is given by the following simple inner product:

$$q_i^o = \langle p, \phi(\alpha_i) \rangle = -x_0 \sin \alpha_i + y_0 \cos \alpha_i, \qquad (1)$$

where $\phi(\alpha_i) = \begin{bmatrix} -\sin \alpha_i & \cos \alpha_i \end{bmatrix}^T$ is the unit vector parallel to the i-th image plane that points anti-clockwise around the circle.

The non-orthogonal, perspective, projection is a more accurate model of traditional cameras and is equivalent to the pin-hole camera model. In this case, the projections of p onto the sensors are no longer inner products. In fact, the location of the projected point, q_i^{no}, is given by

$$q_i^{no} = \frac{f}{r + f - \langle p, \phi_i^{\perp} \rangle} \langle p, \phi_i \rangle, \qquad (2)$$
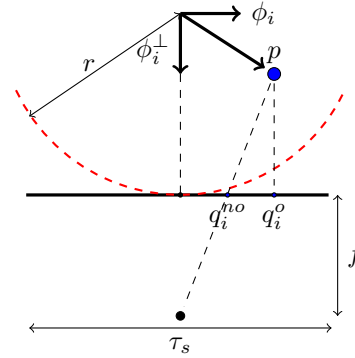

Fig. 3: The orthogonal (q_i^o) and non-orthogonal (q_i^{no}) projection operators applied to a 2-D point p, within a circle of radius r.

where f is the focal length of the camera, $\phi_i = \phi(\alpha_i)$ and $\phi_i^{\perp} = \phi^{\perp}(\alpha_i) = \begin{bmatrix} \cos \alpha_i & \sin \alpha_i \end{bmatrix}^T$ is the unit vector, orthogonal to ϕ_i, that points towards the centre of the i-th image plane. Note that, as $f \to \infty$, $q_i^{no} \to q^o$.

Finally, we will assume that the cameras have equal length image planes, denoted τ_s, chosen so that each camera covers the whole region of interest. I.e. $\tau_s = 2r$ in the orthogonal case and $\tau_s = \frac{2fr}{\sqrt{f^2 + 2fr}}$ in the non-orthogonal case.

2.2. The Image Acquisition Pipeline

The pin-hole camera model is equivalent to a perspective projection; however, a more realistic model, such as the one depicted in Fig. 2, must also account for the camera optics and sampling process.

Let us represent the projection of the point p, onto the image plane of the i-th sensor, as the 1-D function $g_i(u) = \delta(u - q_i)$; i.e., a Dirac at the location q_i. The effect of the camera lens can then be modelled by the following convolution:

$$h_i(u) = g_i(u) * \Psi(u), \qquad (3)$$

where $\Psi(u)$ is the point-spread function [26] which is commonly modeled by the Airy disc.

To model the sampling operation of an imaging sensor we apply a further convolution followed by ideal sampling. For an n-pixel sensor, we choose the sampling kernel, $\Xi(u)$, to be a box function with support $w = \frac{\tau_s}{n}$; i.e., the pixel width. This models the fact that image sensors usually work by integrating all the light rays which fall into each pixel region [27]. Finally, the ideal sampling has period w.

3. POINT LOCALISATION AS A FRAME QUANTISATION PROBLEM

In the previous section, we reviewed a traditional image acquisition model which is applicable to a wide range of imag-
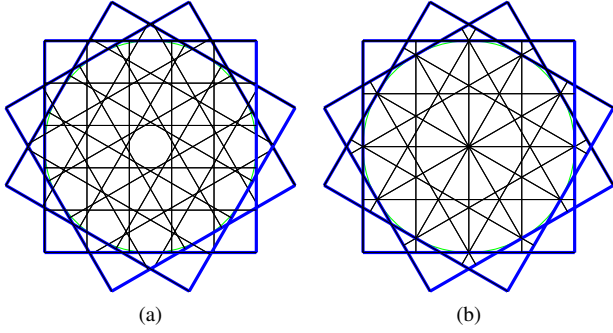
(a)   (b)

Fig. 4: Partitionings of the circle of interest for orthogonal projection with an odd (left) and even (right) number of samples. After sampling, it is impossible to distinguish between two points that lie in the same region.

ing devices, including traditional cameras and computation tomography (CT) devices.

We will now show how this can be interpreted as a quantised frame expansion, which will allow us to derive closed-form expressions for the worst-case localisation error.

To see that this interpretation is valid, note that the camera vectors, $\phi(\alpha_i)$, form a frame in $\mathbb{R}^2$, which we will denote Φ. Therefore, we can consider the projections of the point p onto all the image planes as p projected onto the frame Φ; i.e., $q = \Phi p$.

In addition, if we assume the point activates only a single pixel in each camera[1], the acquisition model, depicted in Fig. 2, is equivalent to the following quantisation:

$$s_i = Q_w(q_i),\qquad (4)$$

where Q_w is the quantisation function defined as[2]

$$Q_w(y) = \left\lfloor \frac{y}{w} \right\rfloor w + \frac{w}{2}. \qquad (5)$$

In quantisation terminology, the pixel width, w, is the quantisation error.

3.1. Orthogonal Projection

We will first investigate if we can localise the point p with infinite precision by using an infinite number of cameras. In the orthogonal projection case, it was shown, in [25], that this is possible iff $\|p\|_2 \geq w$.

This condition can be intuitively understood from Fig. 4: as we add more cameras, there is a central circle which, in the

---

[1]If the PSF causes the point to be spread across multiple pixels, one could exploit more advanced sampling techniques that achieve sub-pixel accuracy.

[2]This definition is valid when n is even. For odd n, $Q_w(y) = \text{sign}(y)\left\lfloor \frac{|y|}{w} + \frac{1}{2}\right\rfloor w.$

case of odd n, is not further divided and, in the case of even n, is only divided through the origin.

To understand this more precisely, let us consider the projection of the point p onto an image plane at angle α as a function of α:

$$q^o(\alpha) = -x_0 \sin\alpha + y_0 \cos\alpha = \|p\|_2 \sin(\alpha + \beta), \qquad (6)$$

where $\beta = \arctan2(y_0, -x_0)$. For even n, the quantised version,

$$s_i = Q_w(q^o(\alpha)) = \left\lfloor \frac{\|p\|_2 \sin(\alpha + \beta)}{w} \right\rfloor w + \frac{w}{2}, \qquad (7)$$

has discontinuities at angles α_{d_i}, where $\|p\|_2 \sin(\alpha_{d_i} + \beta) = k_i w$ for some integer k_i. When $\|p\|_2 < w$, this only occurs at one threshold ($k_i = 0$), which occurs at the two angles $\alpha_{d_1} = -\beta$ and $\alpha_{d_2} = \pi - \beta$. As $m \to \infty$, we are guaranteed to have cameras at these locations; however, since the angles are π radians apart, the image planes are parallel and we can only localise p to be in the central circle of radius w on the line connecting the two camera centres, i.e. on the line with a quantisation threshold of zero.

When $\|p\|_2 \geq w$, there are more than two discontinuity angles and, since $p \in \mathbb{R}^2$, it can be perfectly reconstructed from its projections onto two non-parallel image planes:

$$p = \langle p, \phi(\alpha_{j_1})\rangle \tilde{\phi}(\alpha_{j_1}) + \langle p, \phi(\alpha_{j_2})\rangle \tilde{\phi}(\alpha_{j_2}), \qquad (8)$$

where α_{j_1} and α_{j_2} are the angles of the image planes, $\tilde{\phi}(\alpha_{j_1}) = \frac{1}{\sin(\alpha_{j_2} - \alpha_{j_1})}\begin{bmatrix}\cos\alpha_{j_2} & \sin\alpha_{j_2}\end{bmatrix}^T$ is the dual of $\phi(\alpha_{j_1})$ and $\tilde{\phi}(\alpha_{j_2})$ is the dual of $\phi(\alpha_{j_2})$.

Of course, we can only approximate p because we only have access to quantised versions of these projections:

$$\hat{p} = Q_w(\langle p, \phi(\alpha_{j_1})\rangle)\tilde{\phi}(\alpha_{j_1}) + Q_w(\langle p, \phi(\alpha_{j_2})\rangle)\tilde{\phi}(\alpha_{j_2}).$$

To investigate the accuracy of this approximation, consider two imaginary cameras placed at unknown angles α'_{j_1} and α'_{j_2} such that $Q_w(\langle p, \phi(\alpha_{j_i})\rangle) = \langle p, \phi(\alpha'_{j_i})\rangle$, $i = 1, 2$; i.e., α'_{j_i} is slightly shifted from α_{j_i}, $i = 1, 2$, so that the projection of p is exactly between the two quantisation boundaries.

Using these cameras, we can write the approximated point, $\hat{p}$, as

$$\hat{p} = \langle p, \phi(\alpha'_{j_1})\rangle \tilde{\phi}(\alpha_{j_1}) + \langle p, \phi(\alpha'_{j_2})\rangle \tilde{\phi}(\alpha_{j_2})$$
$$= \left(\tilde{\phi}(\alpha_{j_1})\phi(\alpha'_{j_1})^T + \tilde{\phi}(\alpha_{j_2})\phi(\alpha'_{j_2})^T\right)p. \qquad (9)$$

The localisation error, $\delta = p - \hat{p}$, is thus given by

$$\delta = \left(I - \tilde{\phi}(\alpha_{j_1})\phi(\alpha'_{j_1})^T - \tilde{\phi}(\alpha_{j_2})\phi(\alpha'_{j_2})^T\right)p. \qquad (10)$$

$$\delta = \frac{1}{\sin(\alpha_{j_1} - \alpha_{j_2})}\begin{pmatrix}\cos\alpha_{j_2} & -\cos\alpha_{j_1} \\ \sin\alpha_{j_2} & -\sin\alpha_{j_1}\end{pmatrix}\begin{pmatrix}\sin\alpha_{j_1} - \sin\alpha'_{j_1} & \cos\alpha'_{j_1} - \cos\alpha_{j_1} \\ \sin\alpha_{j_2} - \sin\alpha'_{j_2} & \cos\alpha'_{j_2} - \cos\alpha_{j_2}\end{pmatrix}\begin{bmatrix}x_0 \\ y_0\end{bmatrix}. \qquad (11)$$
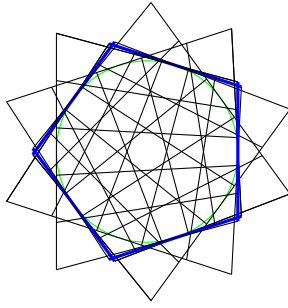
Fig. 5: Partitioning of the circle of interest for non-orthogonal projection with an odd number of samples. After sampling, it is impossible to distinguish between two points that lie in the same region.

Fig. 6: Mean squared error for up to 50 three-sample cameras, in both the orthogonal and non-orthogonal cases.

Substituting the expressions for the image plane vectors and their duals into (10) and applying basic algebraic manipulations yields (11).

By applying $|\sin a - \sin b| \le |a - b|$ and $\left| \alpha_{j_i} - \alpha'_{j_i} \right| \le \frac{2\pi}{m}$ to (11), we can bound the localisation error as

$$\|\delta\|_2^2 \le \frac{4\pi^2 (x_0 + y_0)^2}{m^2} \le \frac{8\pi^2 r^2}{m^2}. \qquad (12)$$

This means $\|p - \hat{p}\|_2^2$ is upper bounded by a term that decreases quadratically with respect to m.

3.2. Non-Orthogonal Projection

Figure 5 shows an example partition for non-orthogonal, perspective, projection with an odd number of pixels. Visually, we can see a similar central circle, which is not subdivided by adding further cameras. In the orthogonal case, the radius of this circle was equal to the pixel width w. In the non-orthogonal case, the radius increases to $\left(1 + \frac{r}{f}\right) w$, which approaches w as $f \to \infty$.

To bound the approximation error, $\delta = \|\hat{p} - p\|_2$, we again analyse the partitionings of quantised frame expansions. In the orthogonal case, the partitioning is created from intersecting rectangles whose shorter edges have length w. In the non-orthogonal case, we have trapezoids, instead of rectangles, with sides of length w and $\left(1 + 2\frac{r}{f}\right) w$.

It follows that each trapezoid has a greater area than the corresponding rectangle with side length w. Therefore, the following bound holds in the non-orthogonal case:

$$\|\delta\|_2^2 \le 8\pi^2 \frac{r^2}{m^2}. \qquad (13)$$

4. SIMULATION RESULTS

In the previous section, we showed how a circular-camera array divides the region of interest into a finite number of partitions. Examples of such partitionings are shown in Figs. 4 and 5[3].

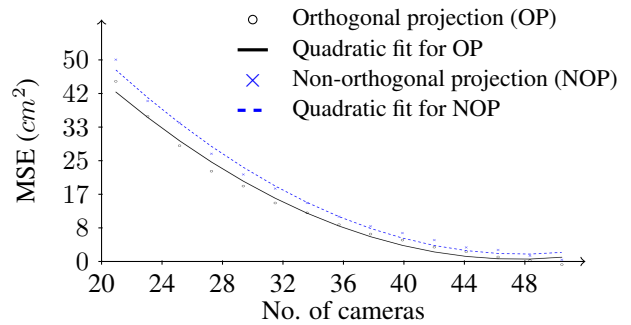[3]You can generate your own partitionings using the web app located at `http://rr.epfl.ch/demos/multicam`.

The interesting semantic behind the constituent regions is that, after projection, we can not distinguish between two points that fall into the same region, meaning there is an inverse relationship between the number of partitions in the circle of interest and the mean squared localisation error.

Figure 6 depicts the mean squared localisation error for a circular-camera array with three-sample sensors and a varying number of cameras. For non-orthogonal projection, a focal length of $f = r$ was used, which leads to a realistic field of view for pinhole cameras. The error values have been well approximated by the reciprocal of a polynomial of degree 2. This suggests that the error decreases quadratically as the number of cameras increases.

Moreover, in Fig. 6, the error decreases slightly faster for orthogonal projection, which corresponds to the theory of Section 3.2. Although this shows a benefit of orthogonal projection, we should note that the minimum sensor size required to cover the region of interest is much smaller when using perspective projection, as we computed in Section 2.1.

5. CONCLUSION

We have reformulated the problem of localising a 3-D point from its images, in a circular camera array, as a frame quantisation problem. This reformulation allows us to derive closed-form worst-case bounds for the localisation error.

We showed that the localisation error can be made arbitrarily low by increasing only the number of cameras. Moreover, we extended the results for orthogonal projections to non-orthogonal projections, which are more common in camera architectures.

In our reformulation, we assumed that the point activates only a single pixel in each camera. We believe that this assumption could be relaxed, by adjusting the quantisation noise model appropriately.

Furthermore, we believe that the frame quantisation interpretation could be used to derive similar results for other multi-camera setups.

6. REFERENCES

[1] J. Y. S. Luh and J. A. Klaasen, "A three-dimensional vision by off-shelf system with multi-cameras," *Pattern*

Analysis and Machine Intelligence, IEEE Transactions on, vol. 7, no. 1, 1985.

[2] H. P. Trivedi, "Can multiple views make up for lack of camera registration?," *Image and Vision Computing*, vol. 6, no. 1, 1988.

[3] H. Aghajan and A. Cavallaro, *Multi-camera networks: principles and applications*, Academic press, 2009.

[4] F. Farshidi, S. Sirouspour, and T. Kirubarajan, "Active multi-camera object recognition in presence of occlusion," in *Proceedings of IROS 2005*. IEEE, 2005, vol. 1.

[5] F. Porikli and A. Divakaran, "Multi-camera calibration, object tracking and query generation," in *Proceedings of ICME 2003*. IEEE, 2003, vol. 1.

[6] A. Ghasemi and M. Vetterli, "Scale-invariant representation of light field images for object recognition and tracking," in *Proceedings of EI 2014*. IS&T/SPIE, 2014, vol. 1.

[7] A. Ghasemi and M. Vetterli, "Detecting planar surface using a light-field camera with application to distinguishing real scenes from printed photos," in *Proceedings of ICASSP 2014*. IEEE, 2014, vol. 1.

[8] A. Ghasemi, N. Afonso, and M. Vetterli, "LCAV-31: a dataset for light field object recognition," in *Proceedings of EI 2014*. IS&T/SPIE, 2014, vol. 1.

[9] T. Georgiev, Z. Yub, A. Lumsdainec, and S. Goma, "Lytro camera technology: theory, algorithms, performance analysis," in *Proceedings of EI 2013*. IS&T/SPIE, 2013, vol. 1.

[10] K. Venkataraman, D. Lelescu, J. Duparr, A. McMahon, G. Molina, P. Chatterjee, and R. Mullis, "PiCam: an ultra-thin high performance monolithic camera array," *Graphics, ACM Transactions on*, vol. 32, no. 6, 2013.

[11] S. Blostein and T. S. Huang, "Error analysis in stereo determination of 3-D point positions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 9, no. 6, 1987.

[12] J. J. Rodriguez and J. K. Aggarwal, "Stochastic analysis of stereo quantization error," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 5, 1990.

[13] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," in *Proceedings of ECCV 2002*, vol. 1. Springer, 2002.

[14] B. Wilburn, V. Vaish, E. V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *Graphics, ACM Transactions on*, vol. 24, no. 3, 2005.

[15] J. X. Chai, X. Tong, S. C. Chan, and H. Y. Shum, "Plenoptic sampling," in *Proceedings of SIGGRAPH 2000*. ACM, 2000, vol. 1.

[16] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," *Computational models of visual processing*, vol. 1, no. 2, 1991.

[17] C. Zhang and T. Chen, "Spectral analysis for sampling image-based rendering data," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 11, 2003.

[18] M. N. Do, D. Marchand-Maillet, and M. Vetterli, "On the bandwidth of the plenoptic function," *Image Processing, IEEE Transactions on*, vol. 21, no. 2, 2012.

[19] C. Gilliam, P. L. Dragotti, and M. Brookes, "On the spectrum of the plenoptic function," *Image Processing, IEEE Transactions on*, vol. 23, no. 2, 2014.

[20] R. Raynor and K. Walli, "Plenoptic camera range finding," in *Proceedings of AIPR Workshops*. IEEE, 2013, vol. 1.

[21] V. K. Goyal, M. Vetterli, and N. T. Nguyen, "Quantized overcomplete expansions in $\mathbb{R}^n$: analysis, synthesis, and algorithms," *Information Theory, IEEE Transactions on*, vol. 44, no. 1, 1998.

[22] B. Beferull-Lozano and A. Ortega, "Efficient quantization for overcomplete expansions in $\mathbb{R}^n$," *Information Theory, IEEE Transactions on*, vol. 49, no. 1, 2003.

[23] S. Rangan and V. K. Goyal, "Recursive consistent estimation with bounded noise," *Information Theory, IEEE Transactions on*, vol. 47, no. 1, 2001.

[24] N. T. Thao and M. Vetterli, "Reduction of the MSE in R-times oversampled A/D conversion $\mathcal{O}(\frac{1}{R})$ to $\mathcal{O}(\frac{1}{R^2})$," *Signal Processing, IEEE Transactions on*, vol. 42, no. 1, 1994.

[25] Z. Cvetkovic, "Source coding with quantized redundant expansions: accuracy and reconstruction," in *Proceedings of DCC 99*. IEEE, 1999, vol. 1.

[26] G. Westheimer and F. W. Campbell, "Light distribution in the image formed by the living human eye," *Journal of the Optical Society of America*, vol. 52, no. 9, 1962.

[27] J. C. Russ, *The image processing handbook*, CRC press, 2010.