Super Resolution for Automated Target Recognition

Maksim Levental

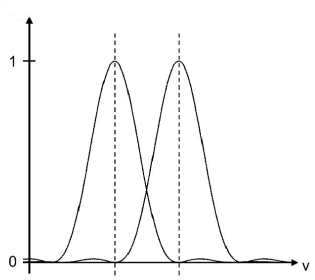

Fig. 1. Rayleigh's criterion[2]

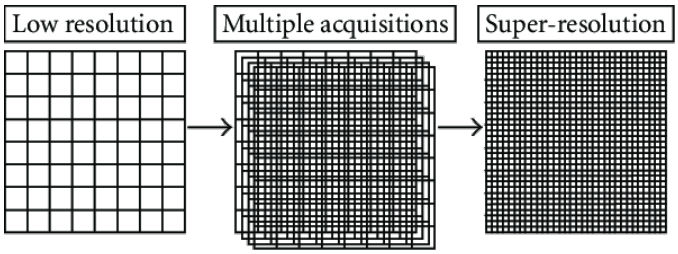

Fig. 2. Multiple image super resolution[3]

***Abstract*—Super resolution is the process of producing high-resolution images from low-resolution images while preserving ground truth about the subject matter of the images and potentially inferring more such truth. Algorithms that successfully carry out such a process are broadly useful in all circumstances where HR imagery is either difficult or impossible to obtain. In particular we look towards super resolving images collected using longwave infrared cameras since high resolution sensors for such cameras do not currently exist. We present an exposition of motivations and concepts of super resolution in general and current techniques, with a qualitative comparison of such techniques. Finally we suggest directions for future research.**

1 Introduction

Super-resolution (SR) is a collection of methods[1] that augment the resolving power of an imaging system. Here, and in the forthcoming, by resolving power we mean the ability of an imaging device to distinguish distinct but proximal objects in the scene. If such objects are modeled as point sources of light then the resolving power of the imaging system is defined by Rayleigh's criterion: two point sources are considered *resolved* when the first diffraction maximum[2] of one point source (at most) coincides with the first minimum of the other (see figure 1).

SR techniques yield high-resolution (HR) images from one or more observed low-resolution (LR) images by restoring lost fine details and reversing degradations produced by imperfect imaging systems. In the case when a single LR source image is used to construct the HR correspondent, the techniques are referred to as single-image-super-resolution (SISR) techniques. These techniques typically operate by either learning some mapping from low resolution chips (uniform partitions of the image, e.g. 3×3 pixels) to higher resolution chips that are highly similar (according to some metrics) and which obey regularity constraints (e.g. agreement at edges). In the case when multiple LR source images are used to construct the single HR correspondent, the techniques are referred to as multiple-image-super-resolution (MISR) techniques. MISR techniques rely on non-redundant and yet pertinent information in multiple images of the same scene (see figure 2). Note that for such information to exist there should be sub-pixel[3] shifts in either the imaging system or the scene between consecutive images.

For typical imaging use-cases, high resolution images are preferable to low resolution images; higher resolutions are desirable in and of themselves and as inputs to later image processing transformations that can degrade image quality (e.g. by virtue of quantization or compression). In theory the resolving power of an imaging system is primarily determined by the number of independent sensor elements that comprise that imaging system (each of which collects a component of the ultimate image). Naturally then, a way to increase the resolution of such a system is to increase the density of such sensor elements per unit area. Unfortunately, and counter-intuitively, since the number of photons incident on each sensor decreases as the sensor shrinks, shot noise[4] thwarts that idea. Furthermore, while sensor density is primary, secondary effects due to optics limit resolution as well; the point spread of a lens (distortion of a point source due to diffraction), chromatic aberrations (distortion due to differing indices of refraction for differing wavelengths of light), and motion blur all function to obscure or erase details from the image.

In domains such as satellite/aerial photography, medical imaging, and facial recognition, high-resolution reconstruction of low-resolution samples is eminently useful since ab-initio

[1]We will often use the verb form "to super resolve" in order to denote the use of one or more such methods.

[2]The amplitude of the diffraction pattern (known as the Airy pattern) of a monochromatic point source through a circular aperture is given by

$$I(\theta) = I_0 \left[\frac{2J_1(ka \sin\theta)}{ka \sin\theta} \right]^2$$

where I_0 is peak intensity (at the center), $k = \frac{2\pi}{\lambda}$ is the wave number of the light, θ is the angle of observation, and J_1 is the Bessel function of the first kind of order one[1]. It is maxima/minima of this function that Rayleigh's criterion concerns.

[3]For example when a point source wholly captured by one sensor element shifts to distributing energy equally amongst the same element and a direct adjacent.

[4]TODO

acquisition of high-resolution images is either logistically difficult or impossible due to aforementioned imaging apparatus limitations. For example in the instance of satellite imagery, acquisition of high-resolution imagery is primarily hampered by optics and physics[5]. In contrast, in the cases of medical imaging (where procedures are invasive and patient exposure time needs to be minimized[6]) and facial recognition (e.g. for purposes of surveillance) the primary challenge is logistics and access to repeat collection opportunities.

The benefits of enhancing images using SR techniques include not only more pleasing or more readily interpretable images for human consumption but higher quality inputs for automated learning systems as well. In particular object detection systems trained on super-resolved images outperform those trained on the low resolution originals[7]. Indeed this is our ultimate goal - not super-resolution per se but super-resolution in the service of improved object detection performance for longwave-infrared (LWIR) imagery. Note that while practically speaking, there exist hardware and software solutions for increasing the resolution of an imaging system, we, owing to a "Ship of Theseus" consideration, discount such propositions. We instead take low resolution images as given and seek techniques that allow for ex post facto reconstruction or inference of precise details. This necessarily constrains techniques under consideration to be algorithmic in nature and software in practice.

The rest of this survey is outlined as follows: Section 2 introduces imaging systems, notation, and the model of imaging that will be the mathematical framework for the proceeding sections, Section 3 surveys classical techniques (those that do not employ neural networks), Section 4 surveys neural-network techniques with heavy emphasis on deep learning (i.e. deep networks), Section 5 discusses the scope and goals of the author's research program, and Section 6 summarizes.

2 Background

2.1 Imaging systems

We begin with a practical discussion of imaging systems. An imaging sensor is a device that converts an optical image into a digital signal. Charge-coupled devices (CCD) and complementary metal-oxide-semiconductor (CMOS) devices are the most common imaging sensors; CCDs have better performance while CMOS devices are newer and less expensive. A third type that's of particular interest to us is the microbolometer, which is used as a sensor in thermal cameras.

CCDs consist of densely packed two-dimensional arrays of buried channel[6] MOS capacitors (see figure 3) with an individual MOS capacitor being the fundamental photon detecting element. Individual MOS capacitors are biased by a gate voltage such that a potential well is produced in the n-type silicon (referred to as the n-channel). This potential well acts as a storage system for charge induced by the inner photoelectric effect[7]. When photons are incident on a MOS capacitor some of the photons are absorbed, some are scattered, and some are transmitted. Those photons that are transmitted interact with electrons in the valence band of the silicon exciting them into the conduction band, and thereby create electron-hole pairs that either diffuse or recombine. For high-quality silicon, the lifetime of such a pair is several milliseconds (before recombination)[10]. The electrons of the electron-hole pairs that do not recombine diffuse into the potential well, while the holes migrate to the grounded substrate (i.e. out of the sensor). Electrons created in this way are called *photoelectrons*.

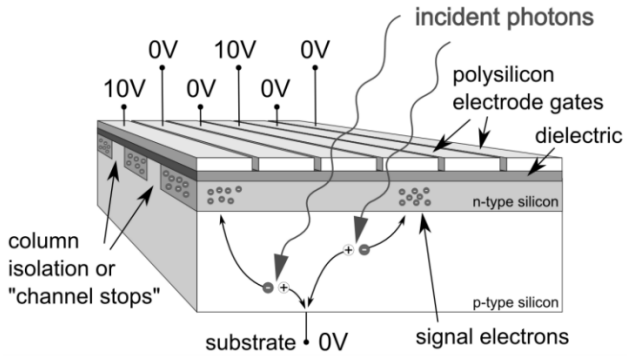

Fig. 3. CCD buried channel MOS capacitor[8]

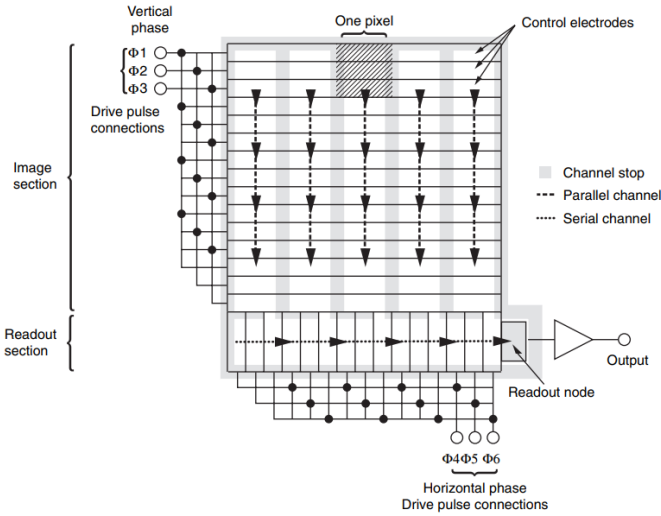

Fig. 4. CCD array[11]

CCD arrays consist of two sub-arrays: an image section and a readout section (see figure 4). The image section is arranged with every third stripe of electrode tied electrically to form three sets of equipotentials. In figure 4 these equipotentials are labeled $\Phi 1, \Phi 2, \Phi 3$, and taken together constitute a vertical register (VR). They function to move the collected

[5]Rayleigh's criterion implies that the angular resolution R of a telescope with optical diameter $D = 2.4$m observing visible light ($\sim$500nm) is approximately[4]

$$R \approx 1.220\frac{\lambda}{D} = 1.220\frac{500\text{nm}}{2.4\text{m}} \approx 0.06\text{arcsec}$$

From an altitude of 250 km this corresponds to a ground sample distance of 6cm. This loss of resolving power is further exacerbated by refraction through turbulent atmosphere[5].

[6]Buried channel as distinct from surface channel. In surface channel MOS capacitors signal charge is stored at the Si-SiO_2 interface, which can lead to charge trapping during the charge transfer process[9].

[7]The photoelectric effect is the emission of electrons when light hits a material. The inner photoelectric effect is that phenomenon but in bulk semiconductors.

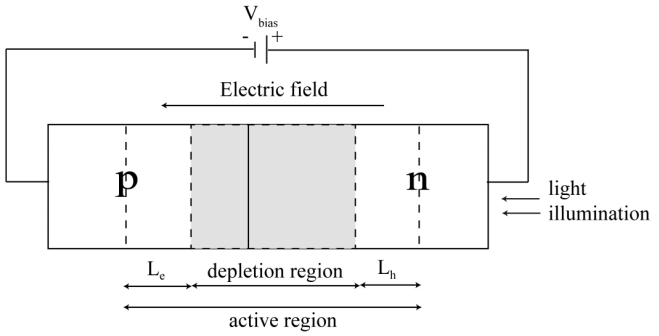

Fig. 5. Photodiode schematic. L_e, L_h are electron, hole diffusion lengths respectively[12]

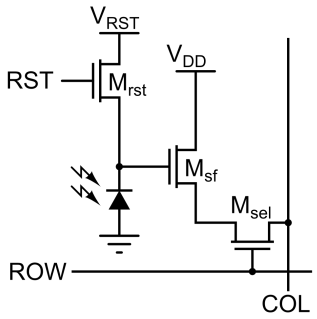

Fig. 6. Three transistor "pixel". M_{rst} is the reset transistor (enabling the photodiode to dump charge), M_{sf} buffers the charge on the photodiode (so that it can be read without loss), and M_{sel} enables a whole row of pixels to be read simultaneously (since all pixels in a physical row are tied to the same row line).

photoelectrons down one electrode line at a time, using charge coupling, while the channel stops function to prevent diffusion of charge across channels. The VR mechanism that shifts collected charge operates as such:

1) Suppose initially there's a collection of photoelectrons on each channel at $\Phi 1$ and only $\Phi 1$. Note this means $\Phi 2, \Phi 3$ are at 0v (again just as in figure 3).
2) $\Phi 2$ is positively biased to 10V. This diffuses the collection of charge under both $\Phi 1$ and $\Phi 2$.
3) $\Phi 1$ is set to 0v. This concentrates the collection of charge under $\Phi 2$.
4) The same is repeated with $\Phi 2, \Phi 3$ and $\Phi 3, \Phi 1$.
5) The entire process repeats thereby shifting the charge three lines (or one pixel row) at a time.

At the bottom of the image section $\Phi 3$ transfers all signal charges to the horizontal register (HR) which functions much like the VR except faster: the HR must transfer every line of pixels independently of all other lines to the read-out node. An obvious challenge faced by this system is how to prevent errant charge from accumulating out of sync with the shift process i.e. how to prevent new photoelectrons from being produced at intermediate lines while far lines are being shifted. The solution is to have interstitial dedicated shift channels in between columns of sensors, with the shift channels being masked off from exposure to light. This type of reading is called *interline transfer* because the accumulated charge is first moved one line over, into the shift channels. Naturally interline transfer shrinks photosensitive area by half and despite possible solutions (e.g micro-lenses being used to focus most of the light onto the unmasked sensors) this is one of the drawbacks of CCDs that CMOS imaging systems do not share.

CMOS sensors consist of arrays of photodiodes (see figure 5). A photodiode is a p-n junction[8] operated in reverse bias mode[9]. When a photon of sufficient energy is absorbed by the diode, it creates an electron-hole pair. If the creation event happens within the active region then the hole moves out through the p-type material and the electron moves out through the n-type material. This establishes a *photocurrent* that can be read by a reading circuit (see figure 6). CMOS sensor arrays do not shift the charge from row to row like CCD arrays. In a CMOS sensor array, each pixel contains a transistor M_{sel} controlled by the voltage applied across a row (see figure 7). In order to read one row of pixels, a row line is raised high to turn on (close) all the M_{sel} transistors in the row. This brings the signals from all the pixels in that row to the shifter register below by way of the column lines. The shift register then outputs the values of the pixels. The high number of transistors needed per pixel in CMOS arrays has only recenty been manageable for semiconductor foundries. This, along with such artifacts as the "rolling shutter" effect produced by rowline reading, are some of the drawbacks of CMOS arrays relative to CCD arrays. Despite this CMOS arrays have become the most common imaging system in consumer goods such as cell phones and digital cameras due to their relatively simple mechanics.

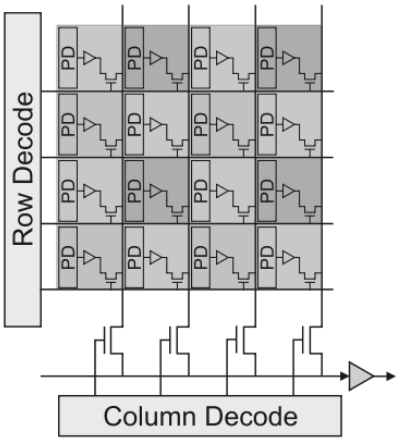

Fig. 7. CMOS array

Both CCD arrays and CMOS arrays only capture visible light. A microbolometer, on the other hand, measures the power in the infrared by exposing a thermistor[10] to the incident light. Since a thermistor's resistance changes as a function of

[8]The interface between a p-type semiconductor (excess holes, i.e. positive charge carriers) and an n-type (excess electrons, i.e. negative charge carriers) semiconductor.

[9]With the p-type material at a lower voltage than the n-type. This causes both the holes and the electrons to flow away from the junction creating a depletion zone.

[10]An element with an electrical resistance that's a function of its temperature.

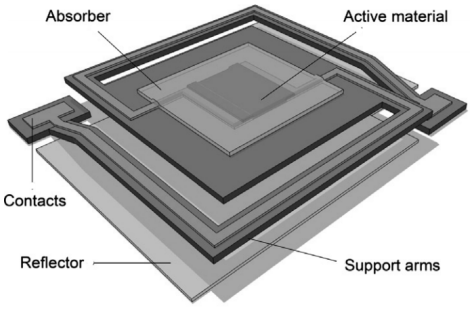

Fig. 8. Bridge structure of Honeywell microbolometer[13]

its temperature, a key issue in the design of a microbolometer is the thermal isolation of the thermistor. With the maturation of micro-machining techniques (such as for MEMS[11] devices) over the last some years, two level microbolometers consisting of a thermo-sensitive component suspended above (and insulated from) silicon have been built (see figure 8). These pixel packages are evacuated and therefore have good conduction, convection, radiation heat transfer properties. The actual thermo-sensitive component consists of a thermistor, an absorber (which aids in transfer of heat to the thermistor), and a reflector that creates a Fabry–Pérot optical cavity[12] (typically $\sim\lambda/4$[14]) that traps the infrared light. Typical materials for the thermistor are vanadium oxide and amorphous silicon owing to their high temperature coefficients of resistance[14], which in effect transform small changes in temperature into large changes in resistance. Measurements of the thermistor are performed by a read-out integrated circuit adjacent to the bridge in the silicon substrate. All told microbolometers are designed much differently from either CCD or CMOS arrays. It is as a result of this fact that high-resolution infrared cameras are not available.

Across all of these imaging systems there are ample avenues for the introductions of the kinds of errors that degrade image quality and across all of these imaging systems there are structures that impose limitations on resolving power. With that in mind we now proceed to formalizing the problem of super-resolution.

2.2 Mathematical notation

Upper case plain latin X, Y denote channel $\times$ row $\times$ column *tensors*[13] representing LR and HR images respectively, with $(0, 0)$ corresponding to the top left corner of the image. Often for the sake of simplicity we consider greyscale images in which case we omit the channel dimension. Lower case plain latin x, y denote LR and HR *patches*[14] respectively. D, H, F, G variously refer to functions that operate on images. Bolded latin $\boldsymbol{X}, \boldsymbol{Y}$ denotes batches.

[11] Micro-electro-mechanical systems.

[12] An optical cavity made from two parallel reflecting surfaces that passes light only when it is in resonance with the cavity.

[13] A multidimensional array[15]. Not to be confused with the algebraic object.

[14] $k \times k$ pixel window, e.g. 3×3.

2.3 Imaging model

Figure 9 shows a conceptual model of the imaging process as carried out by an imaging system. The input to the system is a natural scene that is in effect sampled by the imaging system. In the idealized case the sampling is done at (or above) the Nyquist rate and no aliasing occurs. In practice there is noise and loss introduced at every step of the process: atmospheric turbulence plays a role at large distances, motion produces multiple views of the same scene but also induces blur, imperfections of the lenses further blur the image, and finally down-sampling by the sensor elements into pixels produces aliasing artifacts[15]. The noisy, blurry, down-sampled images are then further degraded by sensor noise. Each such image we call an LR sample.

Let Y denote an idealized HR image of the scene from some fixed vantage point and assume the imaging system collects K LR samples X_k of Y. Formally the X_k are related to Y by

$$X_k = (D_k \circ H_k \circ A_k)(Y) + \epsilon \tag{1}$$

where for the kth sample A_k is the affine transformation representing motion (rigid and perspective shift), H_k represents the composite blur operator (motion and optics blur), D_k represents the down-sampling operator, and ϵ represents the composite noise (environment and sensor noise). The challenge of super-resolution is to solve the inverse problem of finding Y from one or several X_k. In general, since A_k, H_k, D_k are highly degenerate functions, the corresponding inverse problems are ill-posed without regularization and conditioning. The techniques that have been brought to bear on the problem range from interpolation to statistical estimation to example based learning.

3 Classical Algorithms

4 Deep Learning Algorithms

5 Future Research

6 Conclusion

7 Appendix

TODO: work out diffraction circular aperture TODO: work-out poisson noise

Acknowledgments

[15] CCD arrays, for example, employ 2×2 or 3×3 pixel binning, which is the practice of collapsing windows of pixels down to one pixel.

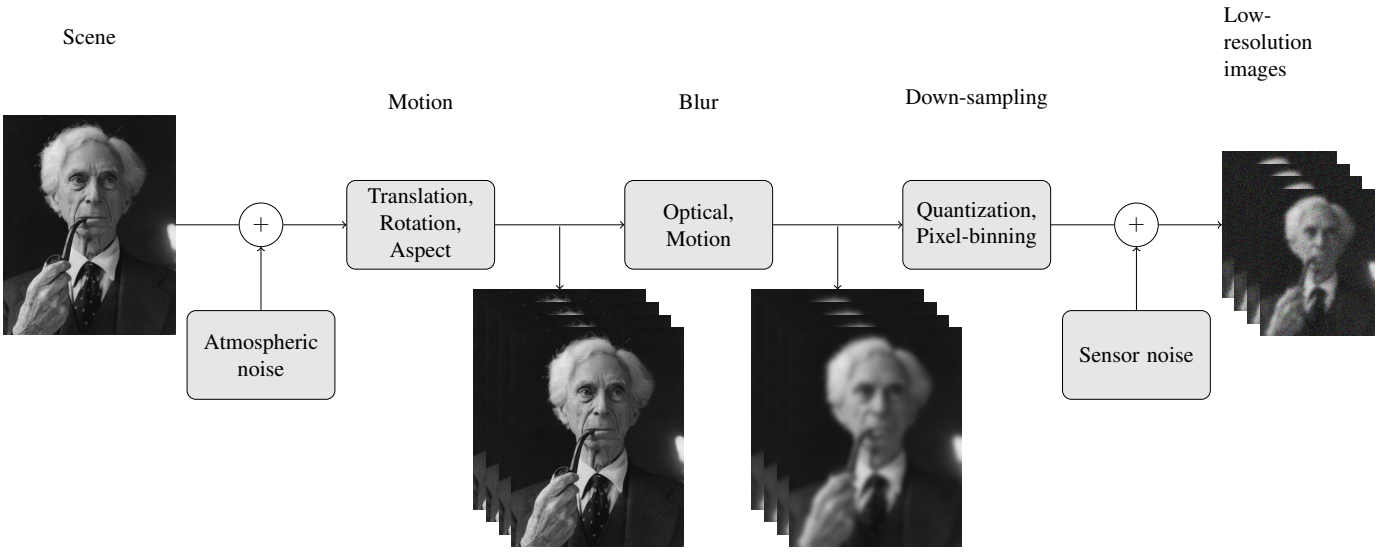

Fig. 9. The imaging model illustrating the relationship between a scene and final low-resolution images due to noise, motion, blur, and sampling.

REFERENCES

[1] J. W. Goodman, *Introduction to Fourier Optics*, 3rd. Roberts & Co. Publishers, 2005, pp. 76–78.

[2] P. Scholz, "Focused ion beam created refractive and diffractive lens techniques for the improvement of optical imaging through silicon," PhD thesis, Jul. 2012. DOI: 10.14279/depositonce-3270.

[3] J. Kennedy, O. Israel, A. Frenkel, R. bar-shalom, and H. Azhari, "Improved image fusion in pet/ct using hybrid image reconstruction and super-resolution," *International journal of biomedical imaging*, vol. 2007, p. 46 846, Jan. 2007. DOI: 10.1155/2007/46846.

[4] L. R. F.R.S., "Xxxi. investigations in optics, with special reference to the spectroscope," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 8, no. 49, pp. 261–274, 1879. DOI: 10.1080/14786447908639684. eprint: https://doi.org/10.1080/14786447908639684. [Online]. Available: https://doi.org/10.1080/14786447908639684.

[5] D. L. Fried, "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures," *J. Opt. Soc. Am.*, vol. 56, no. 10, pp. 1372–1379, Oct. 1966. DOI: 10.1364/JOSA.56.001372. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=josa-56-10-1372.

[6] E. Van Reeth, I. W. K. Tham, C. H. Tan, and C. L. Poh, "Super-resolution in magnetic resonance imaging: A review," *Concepts in Magnetic Resonance Part A*, vol. 40A, no. 6, pp. 306–325, 2012. DOI: 10.1002/cmr.a.21249. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cmr.a.21249. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cmr.a.21249.

[7] J. Shermeyer and A. V. Etten, "The effects of super-resolution on object detection performance in satellite imagery," *CoRR*, vol. abs/1812.04098, 2018. arXiv: 1812.04098. [Online]. Available: http://arxiv.org/abs/1812.04098.

[8] M. Robbins, "Final test guideline," May 2014.

[9] M. Bass, C. DeCusatis, J. Enoch, V. Lakshminarayanan, G. Li, C. Macdonald, V. Mahajan, and E. Van Stryland, *Handbook of Optics, Third Edition Volume I: Geometrical and Physical Optics, Polarized Light, Components and Instruments(Set)*, 3rd ed. New York, NY, USA: McGraw-Hill, Inc., 2010, ISBN: 0071498893, 9780071498890.

[10] J. R. Janesick, T. Elliott, S. Collins, M. M. Blouke, and J. Freeman, " Scientific Charge-Coupled Devices," *Optical Engineering*, vol. 26, no. 8, pp. 692–714, 1987. DOI: 10.1117/12.7974139. [Online]. Available: https://doi.org/10.1117/12.7974139.

[11] J. Pawley, *Handbook of Biological Confocal Microscopy*, ser. Cognition and Language. Springer, 1995, pp. 918–919, ISBN: 9780306448263. [Online]. Available: https://books.google.com/books?id=16Ft5k8RC-AC.

[12] Y. Xu, "Fundamental characteristics of a pinned photodiode cmos pixels," 2015.

[13] Y. E. Kesim, E. Battal, M. Y. Tanrikulu, and A. K. Okyay, "An all-zno microbolometer for infrared imaging," *Infrared Physics & Technology*, vol. 67, pp. 245–249, 2014, ISSN: 1350-4495. DOI: https://doi.org/10.1016/j.infrared.2014.07.023. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1350449514001479.

[14] R. Ambrosio, M. Moreno, J. Mireles Jr., A. Torres, A. Kosarev, and A. Heredia, "An overview of uncooled infrared sensors technology based on amorphous silicon and silicon germanium alloys," *physica status solidi c*, vol. 7, no. 3-4, pp. 1180–1183, 2010. DOI: 10.1002/pssc.200982781. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pssc.200982781. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/pssc.200982781.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016, p. 33, ISBN: 0262035618, 9780262035613.