University of Oulu
Nano and Molecular Systems Research Unit
Physics

VTT Technical Research Centre of Finland
Sensing solutions
Optical measurements

Tuomas Soudunsaari

Stereo vision augmented Medipix3 for real-time material discrimination

Master's Thesis
Oulu, October 8, 2021

Supervisor: Professor Matti Alatalo, University of Oulu
Advisors: Vili Kellokumpu D.Sc. (Tech), VTT
Janne Paaso D.Sc. (Tech), VTT

Abstract

A secure and sustainable supply of raw materials, such as minerals and metals, is a major challenge for the European Union. Domestic consumption of minerals and metals substantially exceeds production, leading to a high reliance on imports to meet demand, which is driven by the various sectors of the EU's economy.

This thesis was conducted as a part of the Horizon 2020 funded X-MINE project, which aims to address the issue by combining novel sensing technologies to improve resource characterization and economic feasibility of domestic mining operations. The focus of the thesis is within early stage ore extraction, where a real-time sensing platform can be employed in a way to reduce waste at an early stage, decreasing the environmental footprint generated by downstream processing.

The advances in CMOS technology have enabled the development of photon counting hybrid pixel detectors, such as Medipix3, for noise-free, high resolution X-ray imaging. Combined with high-speed readout electronics, Medipix3 can be used for imaging and identifying higher density intrusions within ore samples in real-time. In addition, developments in GPU-accelerated stereo vision and related hardware have resulted in high-speed stereo vision camera systems. The combination of stereo vision and Medipix3 is explored in this thesis for real-time material discrimination purposes.

In collaboration with multiple European mines, ore samples with reference elemental data were obtained and measured using the combination of stereo vision and Medipix3. In addition, Medipix3 supports a simultaneous dual channel measurement mode, which is used to form a comparison to the performance of the presented system. Measurement principles and physics for both X-ray imaging and stereo vision are discussed. Additionally, calibration methods and algorithms for both measurement modes are presented. Furthermore, methods for data fusion and algorithm performance evaluation are outlined.

Results for both measurement modes are presented, along with the relevant measurement physics. Superior performance is obtained with the aug-

mentation of stereo vision, in part due to adverse effects of high-speed imaging on image quality with X-ray imaging. Dual channel approach requires higher data throughput, which results in a reduced integration time in comparison. Additionally, charge sharing effects due to high resolution reduce the spectral measurement capabilities.

Foreword

Acknowledgements

First of all, I would like to provide my sincere thanks to Professor Matti Alatalo for supervising this thesis. Furthermore, I am extremely grateful for the help and support from my thesis advisors Vili Kellokumpu, D.Sc., and Janne Paaso, D.Sc., at VTT. I would also like to extend my thanks to Tapio Vaarala, M.Sc., for building the X-ray laboratory at VTT, introducing me to modern X-ray detectors and teaching the various insights involved. Additionally, I would like to extend my appreciation to Pekka Kilpeläinen, M.Sc., and Tuomas Seppälä, M.Sc., at VTT for their hard work on the software, with the goal of realizing the aspects of this thesis towards real world use.

This thesis would not have been possible without the work of various people involved in the X-MINE project. First, I want to thank the people at Advacam Oy/s.r.o. for their hard work on the detector and the high-speed electronics readout. I would like to especially thank Stepan Polansky, M.Sc., for the measurements and help provided. In addition, I would also like to express my gratitude to Jan Jakubek, Ph.D., for the fruitful discussions. I also want to thank the people at Antmicro Ltd, especially Piotr Katarzynski, Ph.D., and Wojciech Gryncewicz, M.Sc., for their hard work on the stereo vision camera system and helping conduct the measurements. In addition, my thanks go to Jakub Progorowicz, M.Sc., and Artur Szewczuk, M.Sc., at Comex Polska Sp. z o.o., for providing the research sorter prototype for measurement purposes and helping conduct the measurements.

Finally, I would like to express my deepest gratitude to my mother and father for their love and support throughout the years.

Funding

This work is part of the X-MINE project. The X-MINE project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No [730270].

Contents

Chapter 1

Introduction

A secure supply of raw materials is essential to the economy of the European Union. Due to the growth of global market, EU's share of global metals production has decreased from 50% in 1850, to below 5% in 2009 [1]. European Union is far from being self-sufficient for many metal ores, relying on imports to close the supply-demand gap in domestic use. Better resource characterization and improving the mineral extraction efficiency can help increase the economical feasibility of domestic mining operations, and help reduce the supply-demand gap.

The X-Mine project [2] is funded from the European Commission's Horizon 2020 programme. The project aims to address these challenges by combining novel sensing technologies based on X-ray fluorescence, X-ray transmission radiography and 3-D vision. During the project, these technologies are applied to both mineral exploration and extraction purposes. The focus of this thesis is within mineral extraction by combining X-ray transmission and 3-D vision. In addition, the performance for sample discrimination is evaluated for various European mines involved in the project. During mineral extraction, a great deal of waste rocks can be produced. Discriminating ore from waste rock at an early mineral extraction stage helps increase the economical feasibility of the mining operation, as waste rock is removed from further processing. This results in a lower use of energy, water and chemicals during downstream processing.

One of the goals of the project is to demonstrate real-time sample discrimination directly at the mine, where samples are fed into a mechanical system consisting of a conveyor belt, sensors and computer controlled fast pneumatic valves to separate the samples to ore and waste fractions using compressed air. This is visualized in Figure 1.1.

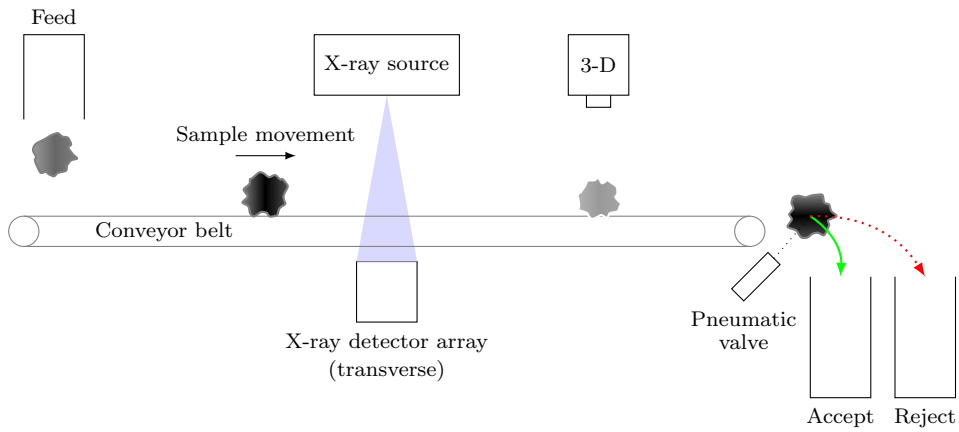


Figure 1.1: Simplified visualization for a mechanical conveyor belt system that combines stereo vision and X-ray attenuation for pneumatic sorting.

As the samples move along the conveyor belt, an X-ray source irradiates the samples and penetrated X-rays are captured by an array of X-ray imaging detectors, producing quantitative information about the samples. The choice of detector here is a high-resolution Medipix3 hybrid pixel detector, benefiting from years of development towards radiation detection by the CERN collaboration[1]. The detector is being commercialized and developed for various applications by Advacam s.r.o. Furthermore, the project consortium converged on a minimum speed requirement of 1.5 m/s for the conveyor belt speed along a detector array width of 420 mm. Combined with the high-resolution imaging data throughput, existing readout electronics for Medipix3 were deemed unsuitable. For this reason, high-speed readout electronics were developed during the project by Advacam s.r.o. Readout electronics speed strictly defines the capacity of the detector as integration time is directly affected, amplifying issues with X-ray photon noise.

X-ray imaging data is augmented by a stereo vision camera system[2] that also images the samples on the conveyor belt. A calibrated dual camera setup produces images that are manipulated in such a way that it enables depth reconstruction. Depth information is then transformed to sample thickness, and used to improve material estimation when single channel mode is used with Medipix3. A bulk of the processing is being done on general purpose graphics processing units, enabling real-time processing speed.

Demonstrations of these technologies are conducted in both research laboratory environments and containerized sorting prototypes. The prototype

[1] `https://medipix.web.cern.ch/`

[2] developed by Antmicro Ltd

units are deployed to mines as a part of a piloting operation in order to demonstrate sorting and its performance in comparison to existing sorting technologies. The work done in this thesis supports the realization of these goals.

This thesis is organized as follows. Chapter 2 describes a well-known method of generating X-ray photons using X-ray tubes, and the mechanisms relevant to the interaction of X-ray photons with matter. Mechanisms, such as photoelectric effect, Compton and Rayleigh scattering are discussed. These mechanisms lead to the attenuation of an X-ray spectrum, which is described to be characteristic to that material. Different materials lead to different attenuation spectra, and this behaviour can be exploited to introduce a measure for material characteristics, such as attenuation coefficient.

Chapter 3 introduces the Medipix3 photon counting pixel detector developed by the CERN collaboration for radiation detection. For the purposes of real time material discrimination, measuring a complete, pixel-specific X-ray spectrum for a given sample is an impractical and a time-consuming process. Medipix3 employs two digital counters for each pixel, which can be configured to measure two different portions of the spectrum. A dual channel response can be generated on a pixel basis, but with limitations, such as charge sharing between adjacent pixels, which is a dominant effect at the spatial resolution of the sensor. Dual channel approach also introduces additional limitations on integration time, which defines the time a pixel has its shutter open and counts X-ray photons. Integration time is reduced compared to the alternative of using a single counter only.

In addition, Chapter 4 introduces the concept and theory of a stereo vision camera, which can be used to measure a thickness profile of a sample and augment the single channel approach. A single channel approach cannot distinguish between a thin, more dense sample, and a thick, less dense sample, and therefore stereo vision provides complementary information. Material discrimination in this sense aims to discriminate between the more and less dense materials. In Chapter 5, we discuss the methods required for combining the data from both sensors. A transformation between image planes is required, and a linear approach is assumed. As a result, both depth and X-ray intensity information are provided on a pixel basis. This leads us to Chapter 6, where two algorithms are discussed as well as a method for their evaluation at varying operating thresholds. Instrumentation, samples and measurements are discussed in Chapter 7. Chapter 8 introduces the results from the measurements, using the discussed method of evaluation, where an upper bound for algorithm performance is introduced. This upper bound is defined by a computed tomography representation of sample elemental content.

1.1 Notation

Matrices are written in bold uppercase letters, such as $\mathbf{M}$. Vectors are denoted with bold lowercase letters, such as $\mathbf{p}$. A column vector of a matrix is denoted with a bold lowercase letter and a subscript. For example, $\mathbf{m}_i$ would describe the ith column vector of matrix $\mathbf{M}$. Points in a world coordinate system are denoted with symbol $\mathbf{p}$, and a subscript is used to differentiate between coordinate systems. Matrix elements are denoted with an italicized symbol and two subscripts, such as m_{ij} to describe a matrix element of $\mathbf{M}$ at row i and column j. A short-hand notation for a transposed matrix inverse is written as

$$\mathbf{M}^{-T} \equiv (\mathbf{M}^{-1})^T = (\mathbf{M}^T)^{-1}. \tag{1.1}$$

Chapter 2

X-ray generation and interactions with matter

This chapter reviews the dominant processes that contribute to X-ray beam attenuation. The generation of X-ray photons using X-ray tubes is also discussed, as well as the resulting shape of the incident spectrum.

2.1 X-ray generation

A common method to generate X-rays is by an X-ray tube, consisting of a vacuum tube fitted with a cathode and an anode. A cathode is a filament which is heated to thermally expel electrons. A high voltage power source (15-150 kV) is connected across the cathode and anode, accelerating the electrons towards the anode. Acceleration causes electrons to gain a maximum kinetic energy of $E_{\text{kin}} = eU$, where U is the acceleration voltage. Considering a voltage of $U = 150$ kV, accelerated electrons reach maximum kinetic energy of $E_{\text{kin}} = 150$ keV.

Once the electrons reach the surface of the anode, they quickly lose their kinetic energy by traveling through matter. Energy loss is provided by two mechanisms, ionization and radiative processes (bremsstrahlung) [3]. Radiative processes are the source of X-rays, and ionization is mostly heating up the anode. Most of the energy is converted to heat at the anode, and a small percentage of the energy is converted to X-rays. Good thermal conductivity and a high melting point are important properties for an anode material. A typical choice for anode material is tungsten, due to its high atomic number ($Z = 74$), atomic density and thermal properties. A high atomic number increases the efficiency of bremsstrahlung production in comparison to low atomic number alternatives.

Figure 2.1 outlines a simplified geometry of an X-ray tube. Electron beam is directed to an area at the anode, which corresponds to the focal spot size of the X-ray tube. As the beam size increases, more area is covered and more current can flow from cathode to anode before thermal limits of the anode are reached. While increasing focal spot size increases the X-ray photon flux, it introduces a negative impact on image quality at the detector, as the image sharpness is reduced. A trade-off between imaging quality and photon flux is needed. Note that due to geometry, the image of the sample is magnified at the detector.

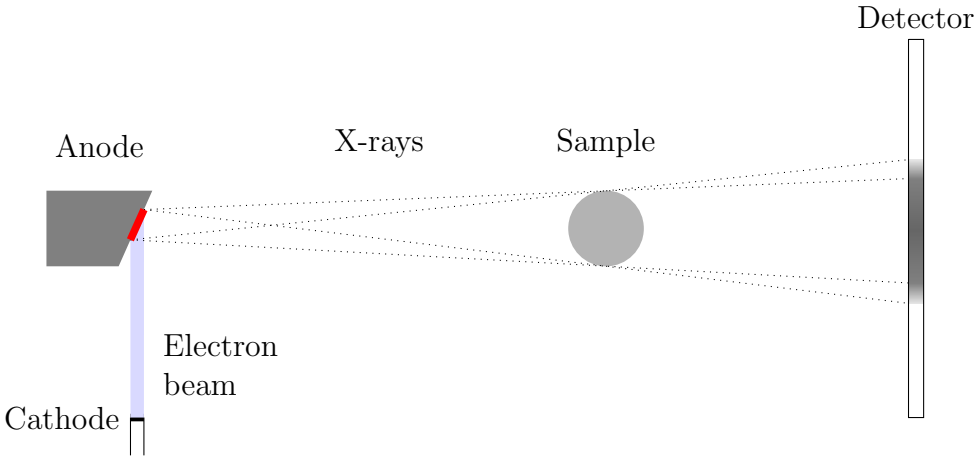


Figure 2.1: X-ray tube. The focal spot is depicted in red.

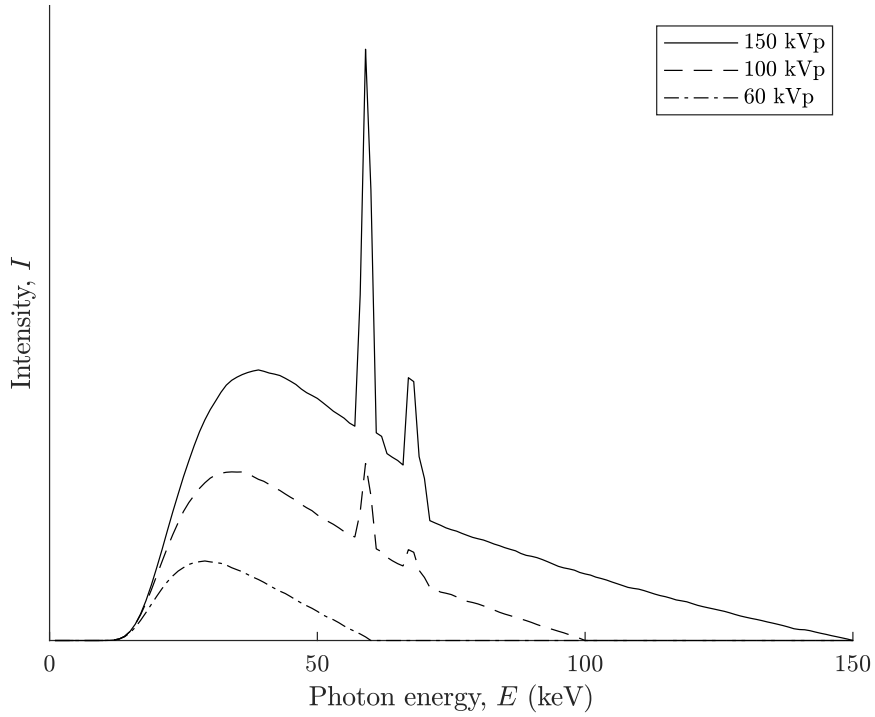


Figure 2.2: X-ray tube spectrum at 150, 100, and 60 kVp, for a tungsten anode.

Figure 2.2 depicts the spectrum of an X-ray tube at various operating peak voltages. There are two components to the spectrum: a continuous component from bremsstrahlung and two peaks from X-ray fluorescence that is due to photoelectric effect, resulting in X-rays being emitted that correspond to characteristic radiation of the tungsten anode. The spectra were simulated using SPEKTR 3.0 toolbox, which enables the X-ray tube output spectrum simulation up to 150 kVp [4].

2.2 X-ray photon-matter interactions

This section focuses on the microscopic description of photon-matter interactions at the photon energy range of 1-150 keV. Interactions are related to macroscopic behaviour by their respective cross sections. Atomic cross sections can be considered as the likelihood of an interaction when an incident beam impinges on an object. Macroscopic behaviour is discussed in Section 2.3.

2.2.1 Photoelectric effect

A photon incident with energy

$$E_\gamma = h\nu \tag{2.1}$$

on a target material can be subject to the process of photoelectric effect if its energy $h\nu$ exceeds the binding energy E_b of a photoelectron in its original shell. The photon is completely absorbed and its energy is transferred to the photoelectron, which is ejected with energy

$$E_e = h\nu - E_b. \tag{2.2}$$

If an inner shell electron is ejected, the atom is in an excited state and the vacancy is filled by an outer shell electron. This de-excitation appears in the form of a characteristic X-ray.

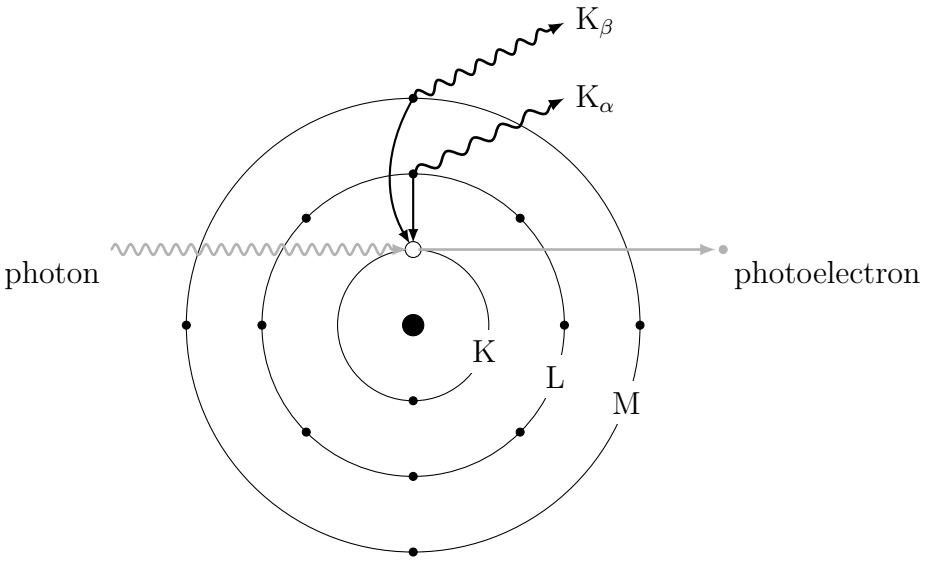


Figure 2.3: Photoelectric effect. An inner shell electron is ejected and filled by an outer shell electron, resulting in a characteristic X-ray (K$_\alpha$ or K$_\beta$, for example) being emitted.

No single analytic expression for the cross section of the photoelectric effect over all ranges of photon energies E_γ and atomic numbers Z has been found, but a rough approximation [5] is

$$\sigma_{\text{pe}} \propto \frac{Z^n}{E_\gamma^{3.5}}, \tag{2.3}$$

where the exponent n varies between 4 and 5, depending on the energy region.

2.2.2 Compton scattering

Compton scattering refers to the inelastic scattering of a photon with an electron that can be considered free relative to the energy of the incident photon. A weakly bound electron satisfies this condition.

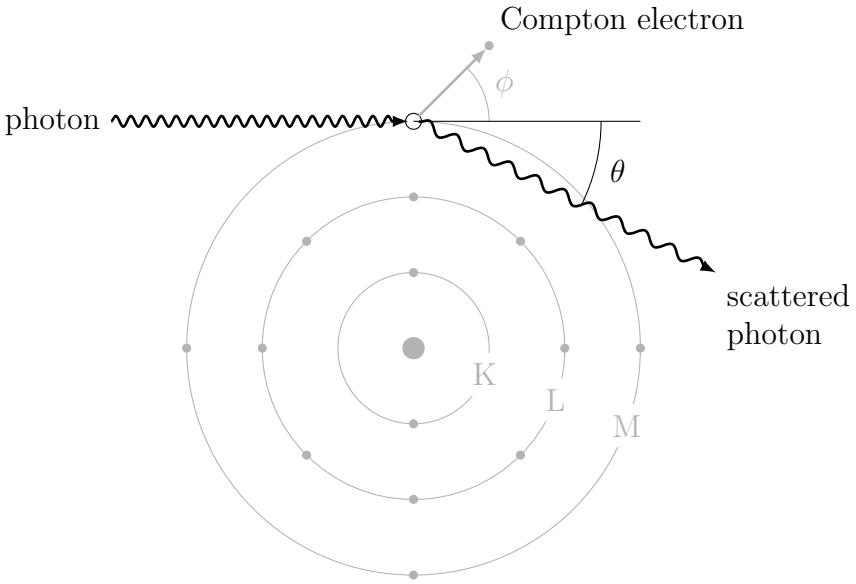


Figure 2.4: Compton scattering.

As a result, a photon transfers some of its initial energy to a Compton electron that is ejected. The photon emerges as a scattered photon with lower energy [6]

$$E' = \hbar\omega' = \frac{\hbar\omega}{1 + \frac{\hbar\omega}{m_e c^2}(1 - \cos\theta)} \tag{2.4}$$

and with an angle θ relative to its original trajectory.

Compton scattering cross section for a single electron σ_c^e can be obtained from the Klein-Nishina formula by integrating the differential cross section over all solid angles [7]. This cross section must be multiplied by the atomic number to obtain the total scattering cross section per atom.

$$\sigma_c = Z\sigma_c^e \tag{2.5}$$

2.2.3 Rayleigh scattering

Rayleigh scattering corresponds to the coherent scattering of a photon from an electron that is tightly bound to an atom. The trajectory of the photon is

slightly changed, but without a change in energy as an elastic process. The derivation is omitted here, but the complex atomic scattering factor yields a dependency of Z^2 for the forward scattering cross section [3]. Additionally, a proportionality to Thomson scattering cross section σ_{Th} is identified, which relates to the elastic scattering of a photon with a free electron and follows from the low-energy limit of Compton scattering. For the coherent forward scattering cross section, we have

$$\sigma_{\text{coh}} \sim \sigma_{\text{Th}} Z^2. \tag{2.6}$$

2.3 Attenuation

2.3.1 Monochromatic beam

Consider a monochromatic beam of intensity I_s from an X-ray source, incident on a sample with thickness x relative to the beam direction. Let us divide the sample to differential slices of thickness dx'. The radiant intensity $dI'(x')$ that emerges from the slice is attenuated relative to the intensity I' that entered the slice by

$$dI' = -\mu I' dx', \tag{2.7}$$

where μ is the linear attenuation coefficient, which is material dependent. By separation of variables, we can solve Equation (2.7) as

$$\int_{I_s}^{I} \frac{dI'}{I'} = \ln(I) - \ln(I_s) = \ln\left(\frac{I}{I_s}\right) = -\int_0^x \mu(x')dx', \tag{2.8}$$

where the latter integration is carried over the depth dimension, and linear attenuation coefficient may change over space as the material can contain other materials. Finally, this yields

$$I(x) = I_s e^{-\int_0^x \mu(x')dx'}, \tag{2.9}$$

which is known as the Beer-Lambert law. We also note that the integral in the exponential of Eq. (2.9) is commutative, implying that the relative order of materials does not matter from a macroscopic perspective. Assuming we can quantize the material to a discrete number of different materials, the commutative property allows us to rearrange the materials and reduces the integral to a sum

$$I(x) = I_s e^{-\sum_{i=1}^{N} \mu_i \Delta x_i}, \quad \text{with} \quad x = \sum_{i=1}^{N} \Delta x_i, \tag{2.10}$$

where μ_i describes the linear attenuation coefficient of the ith material and Δx_i its thickness. For a single material this simplifies to

$$I = I_s e^{-\mu x}. \tag{2.11}$$

A more general form for μ is the mass attenuation coefficient μ_m, which is independent of material density ρ. It is defined by

$$\mu_m = \frac{\mu}{\rho} \tag{2.12}$$

and relates to the photon-matter interactions by their cross sections with the relation

$$\frac{\mu}{\rho} = \frac{N_A}{Au} \sum_i \sigma_i, \tag{2.13}$$

where A is the atomic mass number, N_A is Avogadro's number, u is the atomic mass unit, and σ_i denotes a cross section [3].

Mass attenuation coefficient for a mixture or a compound can be obtained by

$$\mu_m = \sum_i w_i \mu_{m,i}, \tag{2.14}$$

where w_i is the mass fraction of the ith atomic constituent and $\mu_{m,i}$ is the corresponding mass attenuation coefficient.

Figure 2.5 depicts the structure of mass attenuation coefficients as a function of atomic number and photon energy. Discontinuities are due to X-ray absorption edges, corresponding to ionization energies from the inner shells of the atom. As the photon energy exceeds an inner shell ionization energy, mass attenuation coefficient rapidly increases. This is due to photons interacting with the inner shell electrons through photoelectric effect.

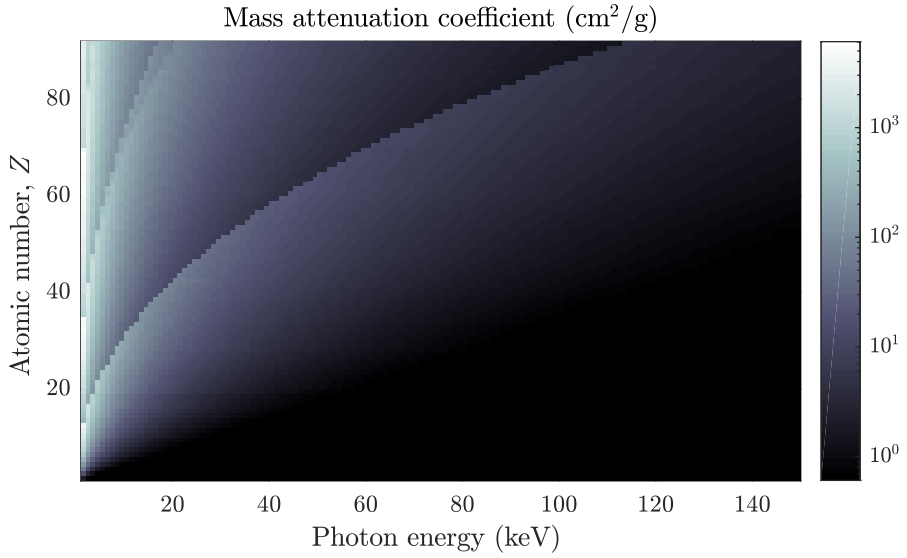


Figure 2.5: Photon mass attenuation coefficient as a function of atomic number and photon energy for the X-ray regime 1-150 keV.

Chapter 3

Medipix3

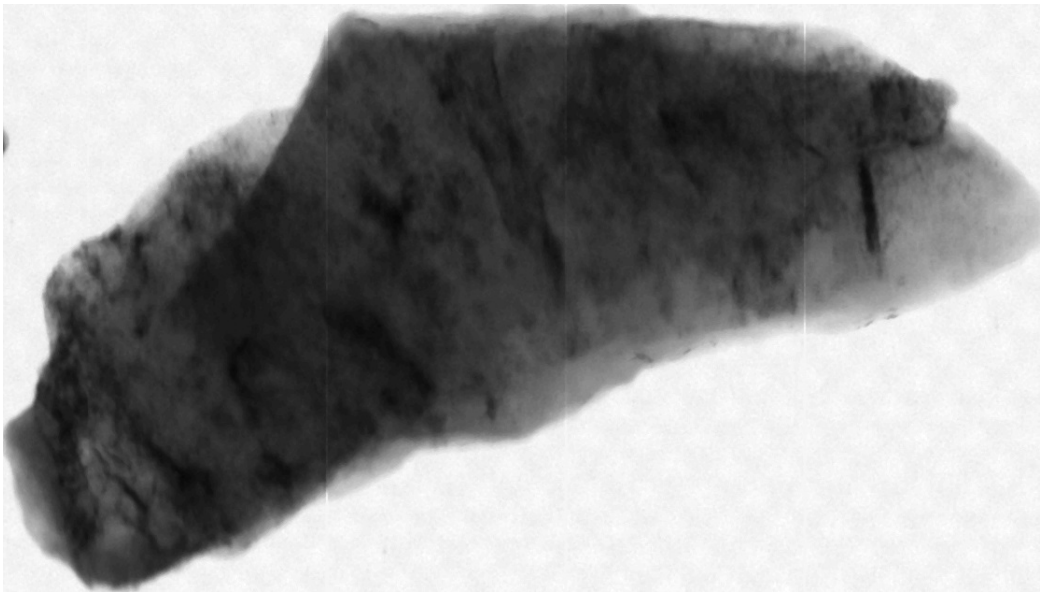


Figure 3.1: Example image (55 x 31 mm^2) of a moving sample produced by Medipix3, using time-delayed integration. Intrusions of higher density material are apparent.

From X-ray generation and photon-matter interactions our focus now transitions to methods of imaging X-rays. History of X-ray imaging starts from the first radiograph ever produced by Wilhelm Röntgen in 1895. Imaging detectors have ever since ranged from photographic plates to modern digitizing X-ray detectors. Many of these are based on energy integrating technologies, where an energy integrating detector (EID) records a signal proportional to the sum of the energy of every photon incident from a source [3]. For a

polychromatic source, information about the energies of individual photons is lost. Additionally, dark current and other sources of noise are present in the images. Most EIDs are indirect conversion systems, where conversion from radiation to a signal happens through an intermediate step provided by a scintillator. A scintillator interacts with the radiation and produces visible light that is measured by a photodetector. Alternatively, direct conversion systems rely on a semiconductor layer that interacts with the radiation by generating electron-hole pairs, which are collected as signal. Photon-counting detectors (PCDs) have the potential to overcome the issues of EIDs by processing photons individually, and mostly through direct conversion. The most common choice for a PCD is a hybrid pixel detector, consisting of a semiconductor sensor layer and a readout chip. The sensor layer and the chip are connected by the process of bump bonding, where a matrix of conductive solder bumps are deposited inbetween, in accordance to the pixel matrix. The development of the first hybrid pixel detector dedicated to medical imaging began in 1997 by the Medipix collaboration. It was preceded by rapid development of hybrid pixel detectors, which was driven by the needs of high energy physics experiments. A prototype detector, Medipix1, consisting of 64x64 pixels with 170 µm pitch was developed [8]. In Medipix1, each pixel is equipped with a single 15-bit counter and a single discriminator for energy thresholding, which can be tuned to remove non-photonic noise. Shortly after the success of Medipix1, the Medipix2 collaboration was formed to build an even more sophisticated chip benefiting from advances in CMOS technology. This resulted in a chip with pixel pitch 55 µm and a matrix of 256x256 pixels [9]. Medipix2 employed two discriminators: one for lower energy threshold and one for higher energy threshold, enabling imaging X-rays for a single energy range between the selected thresholds.

Finally, the Medipix3 collaboration was formed to address issues with charge sharing in Medipix2 [10]. Fine pixelation results in problems with charge transport to its corresponding pixel, where charges are divided to neighbouring pixels and introduce additional counts. Additionally, Medipix3 employed two discriminators and two counters in each pixel, permitting imaging of two energy ranges simultaneously.

3.1 Sensor layer

The choice of detection material for semiconductor sensing layer depends largely on application. Traditionally, silicon (Z_{Si}=14) has been used in high energy physics experiments due to its low mass, reliability and cost [3]. However, silicon is limited by X-ray absorption efficiency at higher X-ray energies.

Alternatively, compound semiconductor media, such as CdTe (Z_{Cd}=48 and Z_{Te}=52) have attracted interest due to high detection efficiency and good room temperature performance. Use of CdTe as a detection medium over the years has been limited by difficulties in producing chemically pure and structurally perfect crystals [11], until recently due to advances of crystal growth and device fabrication technologies. Figure 3.2 depicts the effect of absorption efficiency between silicon and cadmium telluride as a function of photon energy.

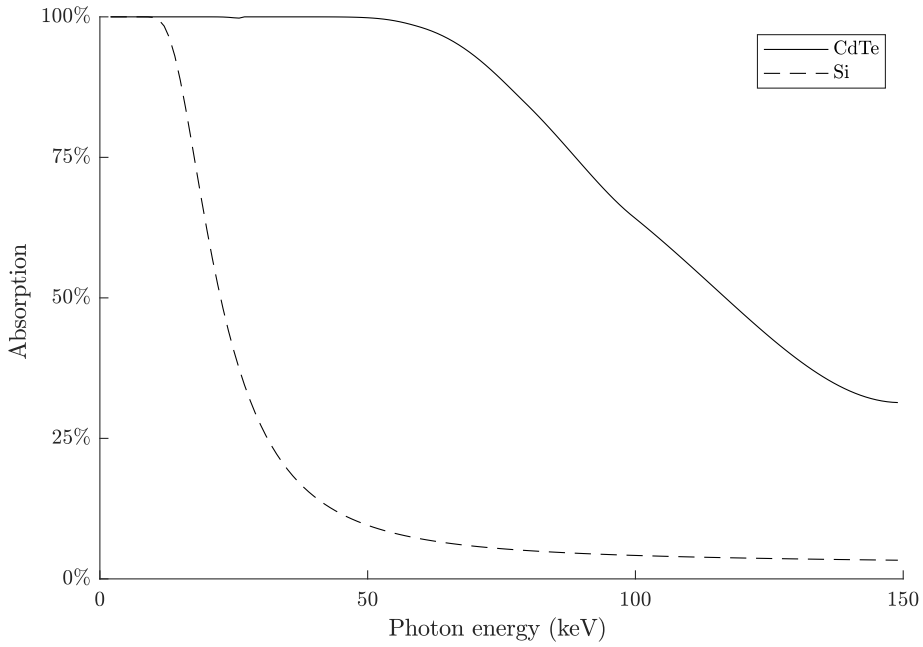


Figure 3.2: The effect of sensor layer visualized for 1 mm CdTe and Si [12]. A sensor layer consisting of CdTe has a higher absorption for hard X-rays relative to a Si sensor layer.

This is partly due to the increased cross section for photon-matter interactions described in Section 2.2 and their dependence on atomic number Z. If the objective is to image materials such as aluminium (or higher Z) with thicknesses on the order of centimeters, lower energies of the X-ray spectrum are easily attenuated by the sample. Therefore, only higher energy photons are left to interact with the sensor layer, and a silicon layer is increasingly inefficient for this purpose. This is also known as beam hardening, where the average photon energy of the spectrum is shifted towards higher energies. Figure 3.3 attempts to visualize the effect of beam hardening, for aluminium

thicknesses of 5 mm and 25 mm.

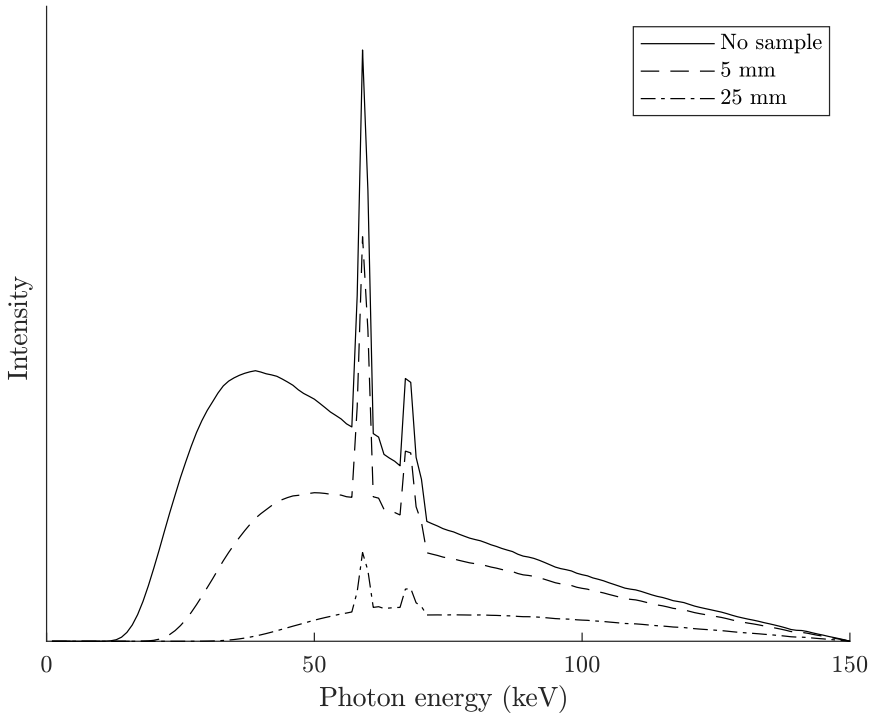


Figure 3.3: Beam hardening due to aluminium thicknesses of 5 mm and 25 mm. Average photon energy is shifted towards higher energies.

Let us continue with the choice of CdTe for detection material due to its absorption efficiency at higher energies. Use of high-Z materials also introduces problems such as fluorescence, which can be no longer neglected. For cadmium telluride, characteristic photon energies peak around 31.8 keV [12], whereas silicon has characteristic photons below 2 keV, which can be filtered using energy thresholding. Another important limitation in some high-Z materials is charge trapping, where semiconductor defects can temporarily trap electrons and holes, reducing the signal induced on the electrodes [13].

3.2 Charge transport

As X-rays interact with the semiconductor layer, a charge cloud consisting of electron-hole pairs is generated. Under the influence of an applied bias voltage, charges drift towards the collection electrodes, where the signal is

amplified and processed by the readout chip. However, due to differences in concentration, charge carriers diffuse during the drift, causing the charge cloud to increase in size [14]. Depending on the depth of the interaction, charge cloud may diffuse laterally to cross pixel boundaries, causing charge sharing between the pixels. Figure 3.4 qualitatively depicts the effect of diffusion across pixel boundaries, at varying depths.

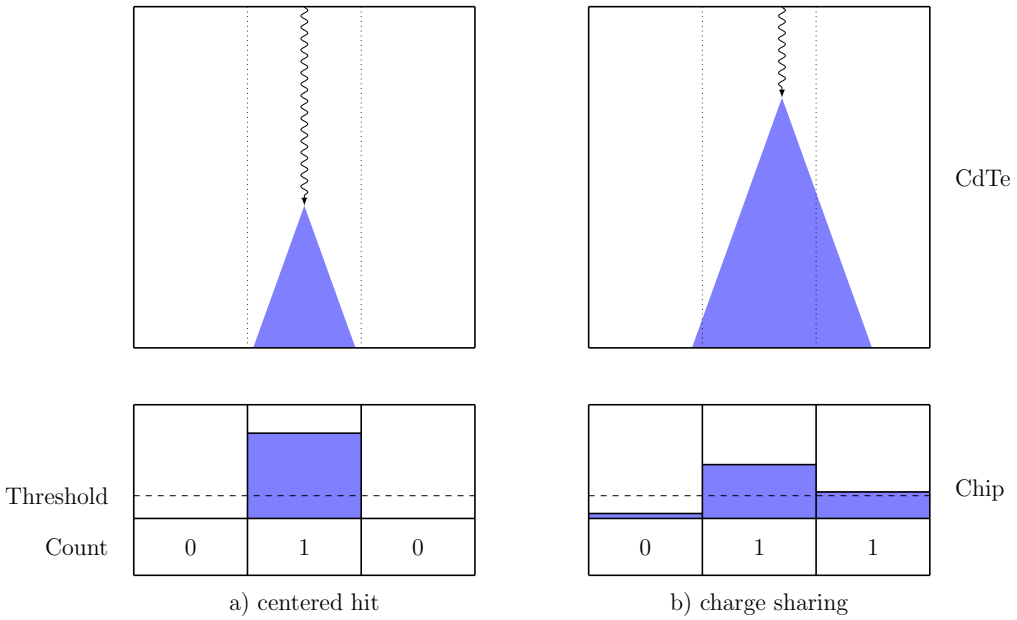


Figure 3.4: a) Centered interaction event with proper photon counting b) Charge sharing introducing false counts to neighbouring pixels and degrading performance

Charges collected at the electrodes are then amplified and compared to a selected threshold, incrementing the counter if threshold is exceeded. Additionally, interaction at a pixel boundary also splits the charge between pixels, where an event may not be counted at all. Figure 3.5 depicts an off-center interaction at a pixel boundary resulting in a missed hit, and a CdTe fluorescence event introducing an additional count.

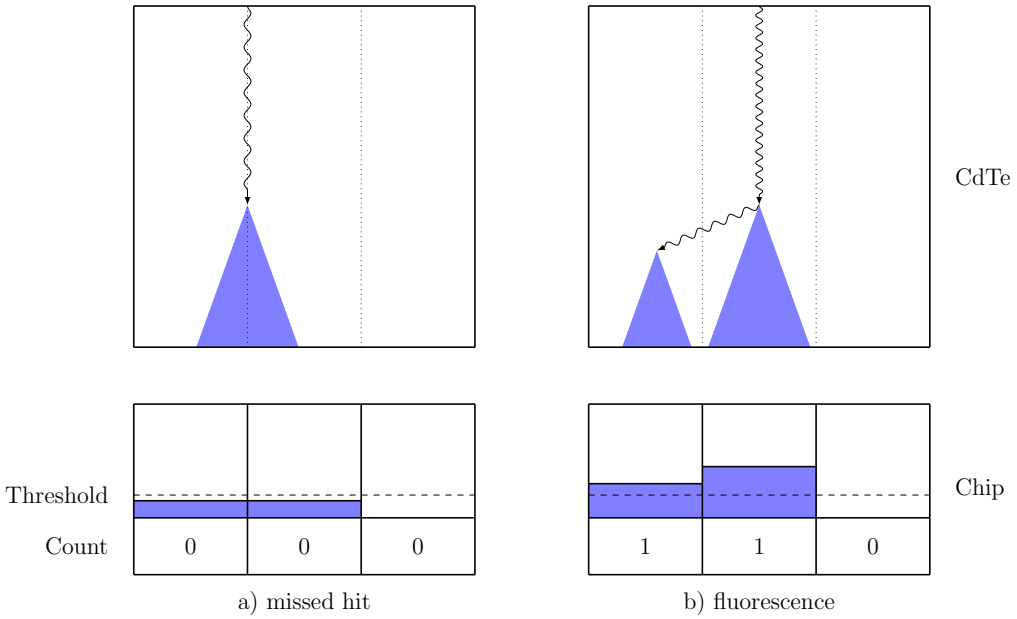


Figure 3.5: a) Off-center interaction event resulting in missed hit b) Centered interaction event with CdTe fluorescence secondary photon introducing additional counts.

3.3 Measurement modes

Medipix3 implements three different measurement modes: single channel mode, dual channel mode and charge summing mode.

3.3.1 Single channel mode

Single channel mode makes use of a single discriminator with counter depth up to 24 bits for a single acquisition. Consider a sample being imaged; for every photon energy, Equation (2.10) applies, and the attenuated intensity interacts with the sensor layer. We let a function $D(E)$ describe the detector efficiency, including absorption efficiency. Therefore, for a monochromatic source, the photon intensity received by the detector is

$$I(E) = I_s(E)D(E)e^{-\sum_{i=1}^{N}\mu_i(E)\Delta x_i} \tag{3.1}$$

For a polychromatic source, the discriminator in each pixel enables us to decide a lower energy threshold for each photon. Intensity measured by the

detector follows

$$I = \int_{E_0}^{E_{\max}} I_s(E) D(E) e^{-\sum_{i=1}^{N} \mu_i(E) \Delta x_i} \, dE, \tag{3.2}$$

where $E_{\max}$ corresponds to the maximum photon energy of the X-ray source, and E_0 is defined by the discriminator. Due to charge sharing, this does not exactly apply, skewing the measurement towards higher photon counts at lower energies. Figure 3.6 visualizes the effect of both sensor layer interaction and the choice of threshold. Photons below the selected threshold value are ignored by the detector.

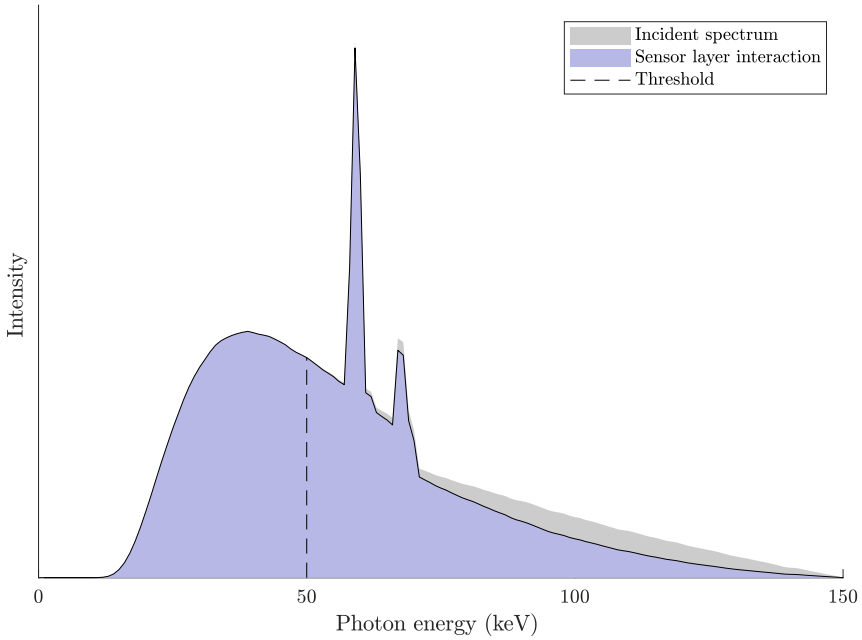
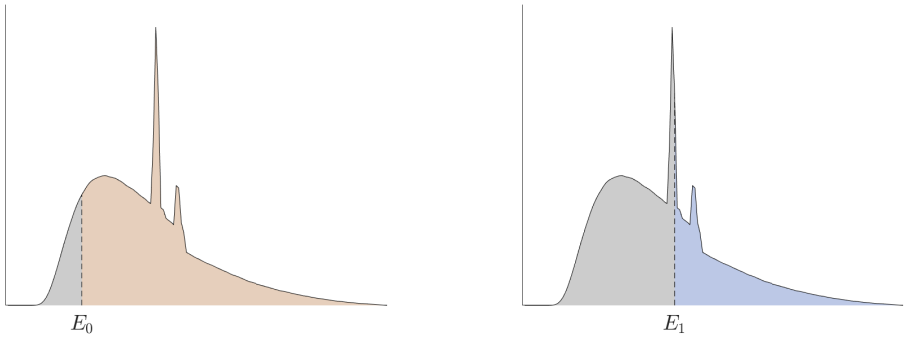


Figure 3.6: Depiction of sensor layer interaction with the incident spectrum and the choice of discriminator threshold. Photons above the threshold are accepted as counts.

3.3.2 Dual channel mode

An additional counter in every pixel means that two different thresholds can be used and two different portions of the incident spectrum can be measured simultaneously, up to 12 bits for each channel. Figure 3.7 visualizes the measured intensities. When an incident photon is measured, the signal goes through both discriminators and independently increments each counter. This means that for a photon which exceeds both thresholds, both counters are incremented.



(a) Threshold 0 (b) Threshold 1

Figure 3.7: Dual channel mode with thresholds E_0 and E_1.

Measured intensities follow

$$I_0 = \int_{E_0}^{E_{\max}} I_s(E)D(E)e^{-\sum_{i=1}^{N}\mu_i(E)\Delta x_i}\,dE \tag{3.3}$$

for the lower channel and

$$I_1 = \int_{E_1}^{E_{\max}} I_s(E)D(E)e^{-\sum_{i=1}^{N}\mu_i(E)\Delta x_i}\,dE \tag{3.4}$$

for the upper channel, where E_0 and E_1 represent selected thresholds.

The benefits of measuring two channels simultaneously become apparent when one discusses noise in the measurements. Measured intensities follow a Poisson distribution, with mean λ and variance λ. From Figure 3.7 it is evident that both channels share a common portion of the spectrum above the higher threshold. One then can recover the intensity between the thresholds by simple substraction, yielding two independent measurements from the spectrum. A difference of two statistically independent Poisson distributions

follows a Skellam distribution[1] with higher variance $\lambda_1 + \lambda_2$ [15], impacting photon statistics. This does not apply here, due to positive noise correlations[2] arising from simultaneous counter increments [16]. After subtraction, the noise from the upper channel is not present in the lower channel. Lower channel simply follows a Poisson distribution with variance equal to the mean, similar to the higher channel.

3.3.3 Charge summing mode

In order to address charge sharing, Medipix3 implements an architecture capable of inter-pixel communication. In single channel or dual channel mode, pixels operate independently. In charge summing mode (CSM), when a charge cloud is generated and laterally diffused across pixel boundaries, pixels employ arbitration logic that compares the pixel outputs and assigns the summed charge to the pixel with the highest collected charge. If the summed charge exceeds the higher threshold, the counter is incremented. Individual pixel outputs below the lower threshold are ignored.

3.4 Flat-field correction

Individual pixels have a different response to the same spectrum input, due to differences in analog circuits integrated within the pixels [17]. A calibration procedure is necessary to produce a uniform response. A simple method is to measure a flood image $\mathbf{F}$ of a flat irradiation field, where no sample is present, with sufficient count rate statistics. With a matrix of individual pixel responses f_{ij}, we can compute the average response $\bar{f} = \frac{1}{n}\sum_{ij}^{n} f_{ij}$ over all the pixels in the matrix. Afterwards, we can compute a matrix of correction coefficients $c_{ij} = \bar{f}/f_{ij}$. After measuring a raw image $\mathbf{R}$ of an unknown sample, individual pixel responses r_{ij} can be corrected to produce a final image $\mathbf{V}$ with responses $v_{ij} = c_{ij}r_{ij}$.

[1] Skellam distribution is useful in describing photon statistics of the difference of two images. One can demonstrate a dual channel response using the single channel mode, with two separate measurements and threshold energies. This setup would yield a Skellam distributed lower energy channel, with variance $\lambda_1 + \lambda_2$. Higher energy channel would still be Poisson distributed, with variance λ_2.

[2] Positive noise correlations convert the lower channel statistics to follow a Poisson distribution with variance λ_1, reducing photon noise in the lower channel. Consequently, both channels are Poisson distributed.

3.5 Signal-to-equivalent thickness calibration

An alternative method to flat-field correction is to calibrate the individual pixel responses as a function of thickness for a known material. Figure 3.8 depicts the intensity attenuation $y(x)$ as a function of sample thickness x. This sample material is usually selected to be representative of the unknown sample to some extent. For example, aluminium is a common choice and it will be used here. Each pixel measures a count rate that is converted to an equivalent thickness relative to this sample material. Therefore, if we measure the calibration sample material, we arrive at the true thickness of the material directly. Due to differences in mass attenuation coefficients, other materials diverge and do not produce accurate thickness estimates.

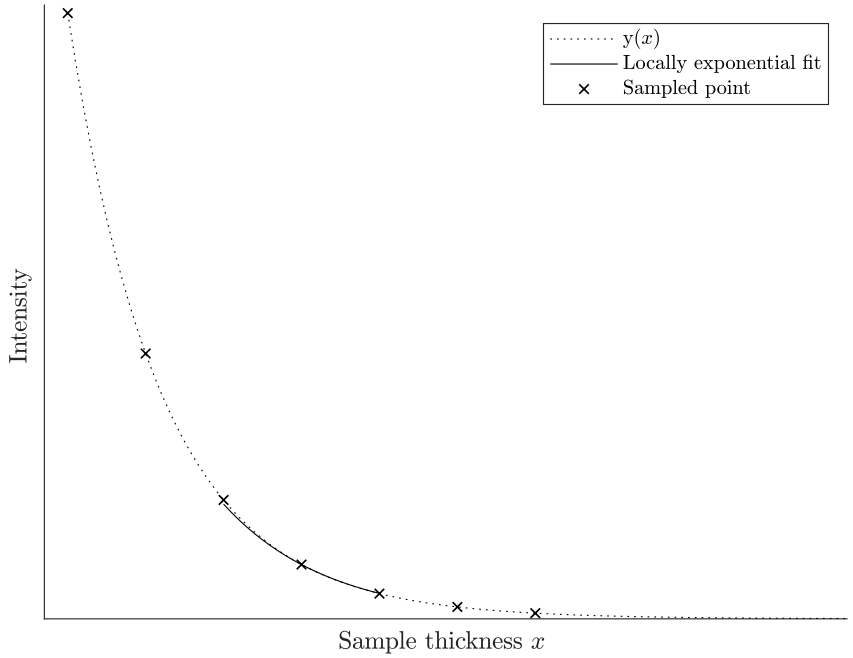


Figure 3.8: Locally exponential fit y_k. The goal is to model the function $y(x)$ based on sampled points. Sampled points are obtained by measuring the intensity of a reference material as a function of different thicknesses. The locally exponential fit attempts to model the function $y(x)$ around the sampled point x_k. The fits around each sampled point are then inverted and used to construct a full function estimate of $x(y)$ using linear interpolation. As a result, subsequent intensity measurements can be transformed to reference material thicknesses directly.

Similar to correction coefficients in flat-field correction, the form of $y(x)$ is specific to a pixel. In practice, the exact form of $y(x)$ is clearly unknown, and the only way to infer its composition is to measure the points at varying thicknesses, at sufficiently small intervals. Once we have sampled the points at thickness x_k, we can approximate $y(x)$ around that point locally by an exponential function with an offset

$$y_k = A_k \exp(a_k x_k) + O_k. \tag{3.5}$$

Fit is done for each calibration point ($k = 1, ..., n$) to arrive at a sequence of parameters A_k, O_k, a_k, for each pixel. After this is completed, we can solve

for x_k

$$x_k = \frac{1}{a_k} \ln \frac{y_k - O_k}{A_k}. \tag{3.6}$$

Naturally, an unknown sample might produce a count rate y, which can be between the calibrated points y_k and y_{k+1} . We can use linear interpolation between the thicknesses x_k and x_{k+1} to arrive at a thickness estimate

$$t(y) = \frac{y - y_{k+1}}{a_k(y_k - y_{k+1})} \ln \frac{y - O_k}{A_k} + \frac{y_k - y}{a_{k+1}(y_k - y_{k+1})} \ln \frac{y - O_{k+1}}{A_{k+1}}, \tag{3.7}$$

assuming the condition $y_k > y > y_{k+1}$ is valid [17]. The result is that we have, to an extent, linearized the response of each pixel, and instead operate in the framework of equivalent thicknesses relative to a calibration material. This serves as a basis for further analysis, especially in the case of dual channel mode.

It is of interest to note that flat-field correction is a weaker method for correcting individual pixel differences. One can attempt to visualize this by plotting multiple pixels as a function of calibration sample thickness, as in Fig. 3.8. Flat-field correction using a flood image simply attempts to normalize the intensity at $x = 0$, but the individual differences in $y(x)$ become apparent as the calibration sample is imaged at varying thicknesses.

3.6 Time-delayed integration

Moving objects tend to introduce motion blur to images. Motion blur can be avoided by matching the integration time of the imaging sensor to the velocity of the object. Thus, the object movement is perceived at an instant of time, where movement is negligible relative to physical pixel size, but at the cost of intensity levels of the image. Time-delayed integration (TDI) is a common technique to capture moving objects while maintaining the intensity levels as one would have with a static image.

In the case of Medipix3, the pixel matrix (256x256) is effectively divided into an array of 256 line scan detectors (1x256). For each line measured, the adjacent line is measured at some time δt later, matching the distance travelled by the object. This procedure is completed for the whole array, and the signal is summed together to increase the total photon count measured. As a result, Medipix3 utilizes the whole pixel matrix as an array of line scan detectors to increase the count rate. An image is still formed, but with a higher count rate than the static imaging counter part, where the integration time of the sensor is matched for the whole matrix with respect to movement.

The value of δt is naturally dependent on movement speed and with simple reasoning we can arrive at the relation

$$\delta t = \frac{h_{\text{eff}}}{v}, \tag{3.8}$$

where h_{eff} is the effective width of the pixel array in the movement direction and v is the movement speed of the sample. Due to geometrical magnification, $h_{\text{eff}} = h/s < 55\,\mu\text{m}$, where h is the pixel pitch ($55\,\mu\text{m}$) and s is the magnification factor. As an example, if the magnification factor is $s = 2$, a pixel array sees a width of $55/2\,\mu\text{m}$, and the velocity at which the camera has to operate is double the sample movement.

Equation (3.8) is a theoretical maximum, and practically unobtainable due to electronics readout time t_r. Therefore,

$$\delta t = \frac{h_{\text{eff}}}{v} - t_r. \tag{3.9}$$

Readout time t_r is specific to a measurement mode, and sets a limit on the operational capacity of the camera with regard to movement speed.

In practice, TDI mode turns both flat-field correction and signal-to-equivalent thickness calibration into a row-based method. In the perspective of measurement pixel statistics, a sum of Poisson distributed variables still follows a Poisson distribution.

Chapter 4

Stereo vision

Stereo vision refers to the ability to infer information about the 3-D structure and distance of a scene, where at least two viewpoints are used to simultaneously image the scene. This 3-D information can be used to augment the X-ray attenuation information as it is evident from Eq. (2.10) that attenuation is a function of sample thickness. Assuming we have two cameras imaging the same scene as Medipix3, we can attempt to reconstruct a depth map containing the measured samples. Furthermore, a background plane corresponding to the conveyor belt surface can be used to transform depth information to sample thickness information.

4.1 Camera calibration

The purpose of camera calibration in machine vision is to calibrate a mapping from physical 3-D coordinates to 2-D image coordinates produced by a camera. This mapping is described by geometric and optical characteristics of a camera (intrinsic parameters), in addition to extrinsic coordinates, which define the position and orientation of the camera relative to a world coordinate system. A set of 3-D coordinates can be generated using a known calibration pattern, such as a checkerboard pattern, preferably with a known grid size in physical units.

4.1.1 Camera model

Let us introduce a planar checkerboard pattern that is imaged by a camera. We define a world coordinate system based on the pattern and its orientation. The inner corners of the pattern can be employed as calibration markers, forming a correspondence between a world coordinate system and an image

coordinate system imaged by a camera.

For a given coordinate $\mathbf{p} = [X, Y, Z]^T$ in the world reference frame, we wish to find a transformation to the coordinate $\mathbf{p}_c = [X_c, Y_c, Z_c]^T$ in a camera reference frame. This can be simply accomplished by a rotation $\mathbf{R}$ followed by a translation $\mathbf{t}$,

$$\mathbf{p}_c = \mathbf{R}\mathbf{p} + \mathbf{t}. \tag{4.1}$$

This defines a set of extrinsic parameters relative to the world coordinate system for a specific camera. The extrinsic parameters consist of orientation and location, with 6 parameters, since the 3-D rotation matrix $\mathbf{R}$ can be decomposed to three Euler angles ω, φ, and κ.

Consider a pinhole camera model as depicted in Fig. 4.1. A point $\mathbf{p}_c$ is projected to a physical image plane π as point $\mathbf{q}$. The image plane is parallel to the xy plane and separated from the projection center $\mathbf{o}$ by focal length f.

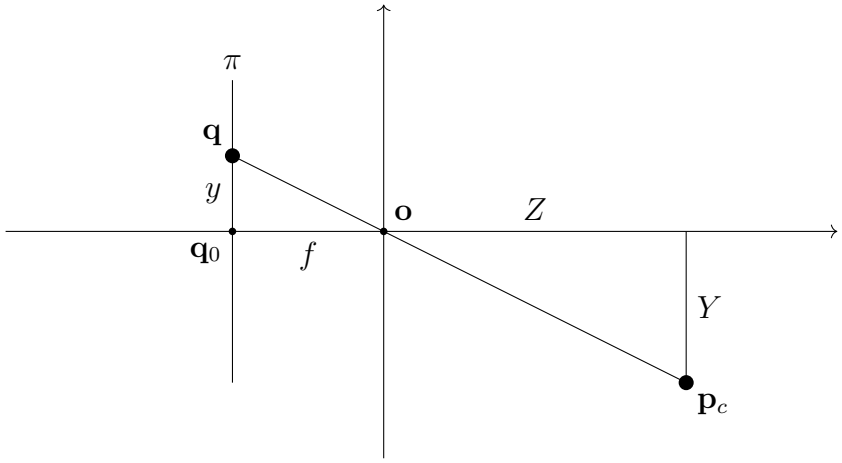


Figure 4.1: Pinhole camera model

From similar triangles, we can see that the y coordinate of the projected point $\mathbf{q}$ follows

$$y = f\frac{Y_c}{Z_c}, \tag{4.2}$$

where Y_c and Z_c are obtained from coordinates of $\mathbf{p}_c$. Extending this to two dimensions, we arrive at the fundamental equations for a perspective camera

$$x = f\frac{X_c}{Z_c} \quad \text{and} \quad y = f\frac{Y_c}{Z_c}. \tag{4.3}$$

The physical image plane π represents an image sensor consisting of a grid of pixels. Digital image coordinates typically start from a corner rather than center, and the camera may have non-square pixels. We can represent this by means of scaling and translation

$$u = \frac{x}{s_x} + u_0 = \frac{f}{s_x}\frac{X_c}{Z_c} + u_0, \tag{4.4}$$

$$v = \frac{y}{s_y} + v_0 = \frac{f}{s_y}\frac{Y_c}{Z_c} + v_0. \tag{4.5}$$

where u and v denote pixel coordinates, s_x and s_y describe axis specific scaling, and $\mathbf{q}_0 = [u_0, v_0]^T$ are the pixel coordinates of the principal point, which ideally corresponds to the image center. These parameters define the set of intrinsic parameters specific to a camera, describing the relation between physical image plane and image coordinates.

With a calibration pattern in its own world coordinate system, we are now nearly equipped to formulate a matrix relation between the world coordinates and image coordinates, representing a linear system of equations with several corresponding points. A problem stands in the form of nonlinearity of $1/Z_c$ in Eqs. (4.4, 4.5) due to perspective projection. This can be avoided by introducing homogeneous coordinates where we extend the points with an additional dimension

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \propto \begin{bmatrix} \lambda u \\ \lambda v \\ \lambda \end{bmatrix} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{4.6}$$

where we can solve for the matrix relation between the coordinates up to a scale factor λ and recover the pixel coordinates of the image plane by dividing the result with λ. Here we have defined $f_x \equiv f/s_x$ and $f_y \equiv f/s_y$ to denote focal depth in pixels. We additionally introduced skewness s for non-orthogonal axes, but for modern cameras this is usually zero. The first matrix contains the intrinsic parameters, which we denote by

$$\mathbf{K} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.7}$$

The use of homogeneous coordinates enables us to express the extrinsic parameters described by Eq. (4.1) in matrix form,

$$\begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix}. \tag{4.8}$$

Given that our selection of a calibration pattern is planar, all the calibration points are coplanar. This lets us choose a world coordinate system where $Z = 0$ for every calibration point, reducing Eq. (4.6) to

$$\begin{bmatrix} \lambda u \\ \lambda v \\ \lambda \end{bmatrix} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \tag{4.9}$$

where we have carried out the matrix multiplication to a single projective transformation matrix $\mathbf{H}$ (or a *homography* [18]) with elements h_{ij}.

The problem is now transformed into finding a solution to $\mathbf{H}$, for which there are various techniques, but a simple one will be presented here. From the calibration pattern we obtain various corresponding points with coordinate pairs $(u_i, v_i) \leftrightarrow (X_i, Y_i, 0)$, where we denote the pair index with i. For a single coordinate pair, the following applies

$$\begin{bmatrix} \lambda u_i \\ \lambda v_i \\ \lambda \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix}. \tag{4.10}$$

We can break down the matrix to a set of equations

$$u_i\lambda = h_{11}X_i + h_{12}Y_i + h_{13}, \tag{4.11}$$
$$v_i\lambda = h_{21}X_i + h_{22}Y_i + h_{23}, \tag{4.12}$$
$$\lambda = h_{31}X_i + h_{32}Y_i + h_{33}. \tag{4.13}$$

Normalizing the homography with constraint $h_{33} = 1$ [19], and eliminating λ gives

$$u_i = \frac{u_i\lambda}{\lambda} = \frac{h_{11}X_i + h_{12}Y_i + h_{13}}{h_{31}X_i + h_{32}Y_i + 1} \tag{4.14}$$

With some algebra we arrive at the relation

$$u_i = h_{11}X_i + h_{12}Y_i + h_{13} - h_{31}u_iX_i - h_{32}u_iY_i. \tag{4.15}$$

Similarly for v_i, we have

$$v_i = h_{21}X_i + h_{22}Y_i + h_{23} - h_{31}v_iX_i - h_{32}v_iY_i. \tag{4.16}$$

Equations (4.15) and (4.16) can be written in matrix form

$$\begin{bmatrix} \vdots \\ u_i \\ v_i \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots & & \vdots & & \vdots & & \vdots & \\ X_i & Y_i & 1 & 0 & 0 & 0 & -u_i X_i & -u_i Y_i \\ 0 & 0 & 0 & X_i & Y_i & 1 & -v_i X_i & -v_i Y_i \\ \vdots & & \vdots & & \vdots & & \vdots & \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} \tag{4.17}$$

or

$$\mathbf{a} = \mathbf{M}\mathbf{h}, \tag{4.18}$$

where $\mathbf{M}$ is a $2n \times 8$ matrix. Matrix $\mathbf{M}$ can be solved using least squares, assuming we have at least $n \geq 4$ corresponding points from the calibration pattern. A linear least squares solution gives us

$$\mathbf{h} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{a}, \tag{4.19}$$

where vector $\mathbf{h}$ contains the matrix elements for constructing homography $\mathbf{H}$.

4.1.2 Zhang's method

With homography $\mathbf{H}$, we can now move on towards estimating both intrinsic and extrinsic parameters using a coplanar calibration pattern such as a checkerboard pattern. Zhang's method [20] is a widely used technique for estimating the parameters, among various other methods not mentioned here. The method provides a closed-form solution for parameter estimation which will be used to demonstrate the basic principles. However, parameters estimated are not optimal, since the closed-form solution attempts to minimize algebraic distance, which is not physically meaningful [20]. Nevertheless, the estimated parameters serve as a good initial guess, which can be refined using nonlinear optimization to provide better results, together with estimating nonlinear effects such as distortion, which are briefly discussed in the next section. The original paper addresses these problems and provides the methods for parameter optimization based on maximum likelihood estimation.

We continue with the homography $\mathbf{H}$ from previous section

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} = \lambda \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}, \tag{4.20}$$

where $\mathbf{h}_i, \mathbf{r}_i, \mathbf{t}$ denote respective matrix column vectors. Multiplying by $\mathbf{K}^{-1}$, we have

$$\mathbf{K}^{-1}\begin{bmatrix}\mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3\end{bmatrix} = \lambda\begin{bmatrix}\mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t}\end{bmatrix} \tag{4.21}$$

Using the orthonormality of $\mathbf{r}_1$ and $\mathbf{r}_2$, we have

$$0 = \mathbf{r}_1^T\mathbf{r}_2 = (\mathbf{K}^{-1}\mathbf{h}_1)^T(\mathbf{K}^{-1}\mathbf{h}_2) = \mathbf{h}_1^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_2 \tag{4.22}$$

and similarly $\mathbf{r}_1^T\mathbf{r}_1 = \mathbf{r}_2^T\mathbf{r}_2$ gives

$$\mathbf{h}_1^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_1 = \mathbf{h}_2^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_2. \tag{4.23}$$

A single homography has 8 degrees of freedom, which is not enough to estimate both extrinsic and intrinsic parameters. There are 6 extrinsic parameters and 5 intrinsic parameters to be estimated. Equations (4.22) and (4.23) give two constraints on the intrinsic parameters for each homography. Therefore, at least three images of the planar calibration pattern from different relative orientations are required. These images produce three different homographies in order to have enough constraints for a solution. Alternatively, if one can assume skewness s to be zero, two images are sufficient.

Let

$$\mathbf{B} = \mathbf{K}^{-T}\mathbf{K}^{-1} = \begin{bmatrix}B_{11} & B_{12} & B_{13}\\ B_{12} & B_{22} & B_{23}\\ B_{13} & B_{23} & B_{33}\end{bmatrix}. \tag{4.24}$$

Matrix $\mathbf{B}$ is symmetric and can be represented as a vector

$$\mathbf{b} = \begin{bmatrix}B_{11} & B_{12} & B_{22} & B_{13} & B_{23} & B_{33}\end{bmatrix}^T. \tag{4.25}$$

Denoting the ith column of $\mathbf{H}$ as $\mathbf{h}_i$, and explicitly writing out the matrix multiplication gives

$$\begin{aligned}\mathbf{h}_i^T(\mathbf{K}^{-T}\mathbf{K}^{-1})\mathbf{h}_j &= \mathbf{h}_i^T\mathbf{B}\mathbf{h}_j && (4.26)\\ &= B_{11}(h_{i1}h_{j1}) + B_{12}(h_{i1}h_{j2} + h_{i2}h_{j1}) && (4.27)\\ &+ B_{22}(h_{i2}h_{j2}) + B_{13}(h_{i1}h_{j3} + h_{i3}h_{j1}) && (4.28)\\ &+ B_{23}(h_{i2}h_{j3} + h_{i3}h_{j2}) + B_{33}(h_{i3}h_{j3}) && (4.29)\\ &= \mathbf{v}_{ij}^T\mathbf{b}, && (4.30)\end{aligned}$$

with

$$\mathbf{v}_{ij} = \begin{bmatrix}h_{i1}h_{j1}\\ h_{i1}h_{j2} + h_{i2}h_{j1}\\ h_{i2}h_{j2}\\ h_{i1}h_{j3} + h_{i3}h_{j1}\\ h_{i2}h_{j3} + h_{i3}h_{j2}\\ h_{i3}h_{j3}\end{bmatrix}. \tag{4.31}$$

Considering the constraints described by Equations (4.22) and (4.23), we can now express them in the form of a homogeneous linear system

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{11} - \mathbf{v}_{22})^T \end{bmatrix} \mathbf{b} = 0, \tag{4.32}$$

or

$$\mathbf{V}\mathbf{b} = 0, \tag{4.33}$$

where $\mathbf{V}$ is a $2n \times 6$ matrix. Solution to Eq. (4.33) can be computed using singular value decomposition and selecting the right singular vector of $\mathbf{V}$ associated with the smallest singular value as the solution to $\mathbf{b}$. However, matrix $\mathbf{B}$ can be constructed using elements of $\mathbf{b}$ only up to a scale factor λ.

Recall the definition of $\mathbf{K}$ in Eq. (4.7). We can write explicitly the elements of $\mathbf{B}$ in terms of intrinsic parameters as

$$\begin{aligned} \mathbf{B} &= \lambda \mathbf{K}^{-T}\mathbf{K}^{-1} \\ &= \lambda \begin{bmatrix} \frac{1}{f_x^2} & \frac{-s}{f_x^2 f_y} & \frac{v_0 s - u_0 f_y}{f_x^2 f_y} \\ \frac{-s}{f_x^2 f_y} & \frac{s^2}{f_x^2 f_y^2} + \frac{1}{f_y^2} & -\frac{s(v_0 s - u_0 f_y)}{f_x^2 f_y^2} - \frac{v_0}{f_y^2} \\ \frac{v_0 s - u_0 f_y}{f_x^2 f_y} & -\frac{s(v_0 s - u_0 f_y)}{f_x^2 f_y^2} - \frac{v_0}{f_y^2} & \frac{(v_0 s - u_0 f_y)^2}{f_x^2 f_y^2} + \frac{v_0^2}{f_y^2} + 1 \end{bmatrix}, \end{aligned}$$

from which intrinsic parameters can be computed as

$$\begin{aligned} v_0 &= (B_{12}B_{13} - B_{11}B_{23})/(B_{11}B_{22} - B_{12}^2) \\ \lambda &= B_{33} - [B_{13}^2 + v_0(B_{12}B_{13} - B_{11}B_{23})]/B_{11} \\ f_x &= \sqrt{\lambda / B_{11}} \\ f_y &= \sqrt{\lambda B_{11}/(B_{11}B_{22} - B_{12}^2)} \\ s &= -B_{12} f_x^2 f_y / \lambda \\ u_0 &= s v_0 / f_y - B_{13} f_x^2 / \lambda. \end{aligned}$$

Using the intrinsic parameters to construct $\mathbf{K}$, the extrinsic parameters for each image are computed with

$$\begin{aligned} \mathbf{r}_1 &= \lambda \mathbf{K}^{-1}\mathbf{h}_1 \\ \mathbf{r}_2 &= \lambda \mathbf{K}^{-1}\mathbf{h}_2 \\ \mathbf{r}_3 &= \mathbf{r}_1 \times \mathbf{r}_2 \\ \mathbf{t} &= \lambda \mathbf{K}^{-1}\mathbf{h}_3 \end{aligned}$$

where $\lambda = 1/\|\mathbf{K}^{-1}\mathbf{h}_1\| = 1/\|\mathbf{K}^{-1}\mathbf{h}_2\|$.

4.1.3 Distortion

In practice, cameras with optics tend to produce physical distortions when imaging objects. If a checkerboard pattern is imaged, grid lines might appear nonlinear in the image. This nonlinearity is usually assumed to be predominantly radial and additionally, in some cases, somewhat tangential. A radial distortion aims to model the distortion caused by a lens system, usually composed of several optical elements. A tangential distortion may be caused by optical elements not being stricly collinear [21]. Distortion effects are qualitatively depicted in Fig. 4.2. Denoting distorted pixel coordinates as $\tilde{u}$ and $\tilde{v}$, displacements can be modeled as $u = \tilde{u} - \delta u^{(r)}$ A radial displacement can be approximated as

$$\delta u^{(r)} = k_1 r^2 + k_2 r^4 + ... \tag{4.34}$$

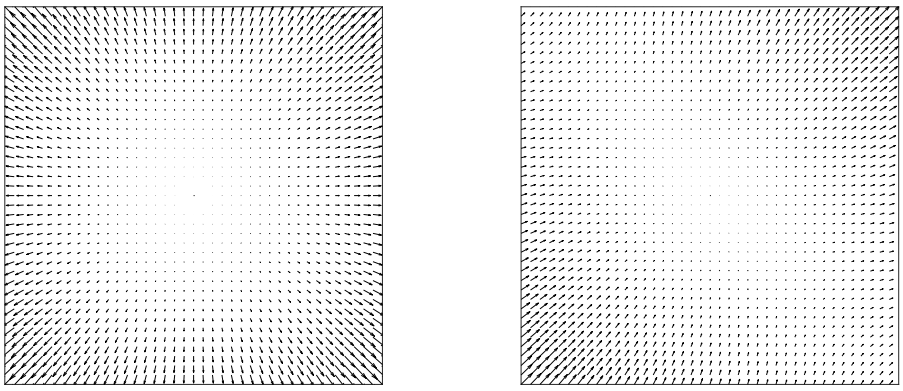
$$\delta v^{(r)} = k_1 r^2 + k_2 r^4 + ..., \tag{4.35}$$

where k_1 and k_2 are coefficients for radial distortion and $r = \sqrt{\tilde{u}^2 + \tilde{v}^2}$. Displacement due to tangential distortion can be approximated as

$$\delta u^{(t)} = 2p_1\tilde{u}\tilde{v} + p_2(r^2 + 2\tilde{u}^2), \tag{4.36}$$

$$\delta v^{(t)} = p_1(r^2 + 2\tilde{v}^2) + 2p_2\tilde{u}\tilde{v}, \tag{4.37}$$

where p_1 and p_2 are coefficients for tangential distortion [19].



(a) Radial displacement (b) Tangential displacement

Figure 4.2: Example of displacement due to a) radial and b) tangential distortion.

4.2 Epipolar geometry

Consider a 3-D point $\mathbf{p}$ in the scene demonstrated by Fig. 4.3 and the objective is to determine its distance relative to a camera. The scene consists of two calibrated cameras, that have been calibrated using the methods mentioned in earlier sections. The cameras are referred to left and right from now on, corresponding to subscript notation l and r. If we consider the left image plane π_l, the point $\mathbf{p}$ is seen as $\mathbf{p}_l$ in the left camera reference frame, and projected as $\mathbf{q}_l$ to the image plane π_l. Determining distance using only one camera here is ambiguous as the point $\mathbf{p}_l$ can lie anywhere on the line described by points $\mathbf{o}_l$ and $\mathbf{p}$.

Let us introduce another camera from different orientation (right-hand side of Fig. 4.3). The line from $\mathbf{o}_l$ towards $\mathbf{p}$ can be projected to the right image plane π_r. This projected line is also known as an epipolar line. The epipolar line contains the point $\mathbf{p}$ projected as $\mathbf{q}_r$ on the right image plane. Point $\mathbf{q}_r$ provides us with information about how to resolve the ambiguity of depth, as we see where on the epipolar line the point stands.

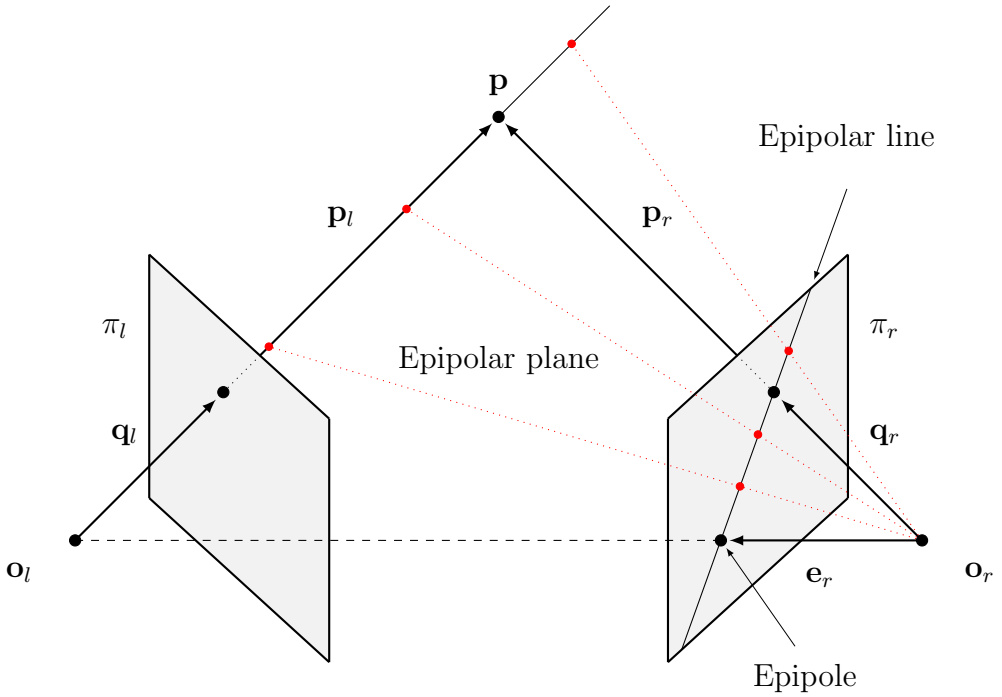


Figure 4.3: Epipolar geometry of a stereo camera system. Search for corresponding point in the right image plane is restricted to the line defined by the right epipole $\mathbf{e}_r$ and projected point $\mathbf{q}_r$. Red lines attempt to emphasize the ambiguity in depth without the addition of a secondary camera, and the need to properly identify corresponding points for accuracy.

We can also consider the reverse case of projecting the right-hand epipolar line to the left image plane. Additionally, we can identify a line from the projection centers $\mathbf{o}_l$ to $\mathbf{o}_r$. This is called the baseline, and the intersection of the baseline with an image plane provides us a point called epipole. Epipole is contained in every epipolar line [22]. Points $\mathbf{o}_r$, $\mathbf{o}_l$ and $\mathbf{p}$ define an epipolar plane.

Points on the left camera frame are related to the right camera frame by the relation

$$\mathbf{p}_r = \mathbf{R}(\mathbf{p}_l - \mathbf{t}), \tag{4.38}$$

where $\mathbf{R}$ describes a rotation matrix and $\mathbf{t} = \mathbf{o}_r - \mathbf{o}_l$ a translation vector between the camera frames. Consider the point $\mathbf{p}$ being visible to both cameras. From camera calibration we know that this point is related to the camera frames by their respective extrinsic parameters as

$$\mathbf{p}_r = \mathbf{R}_r\mathbf{p} + \mathbf{t}_r \tag{4.39}$$

$$\mathbf{p}_l = \mathbf{R}_l\mathbf{p} + \mathbf{t}_l. \tag{4.40}$$

By inverting the extrinsic transformations, we can describe the point $\mathbf{p}$ alternatively as

$$\mathbf{p} = \mathbf{R}_r^T(\mathbf{p}_r - \mathbf{t}_r) = \mathbf{R}_l^T(\mathbf{p}_l - \mathbf{t}_l). \tag{4.41}$$

Therefore,

$$\begin{aligned}\mathbf{p}_r &= \mathbf{R}(\mathbf{p}_l - \mathbf{t}) = \mathbf{R}_r\mathbf{p} + \mathbf{t}_r \\ &= \mathbf{R}_r\mathbf{R}_l^T(\mathbf{p}_l - \mathbf{t}_l) + \mathbf{t}_r \\ &= \mathbf{R}_r\mathbf{R}_l^T(\mathbf{p}_l - \mathbf{t}_l + \mathbf{R}_l\mathbf{R}_r^T\mathbf{t}_r),\end{aligned}$$

or

$$\mathbf{R}(\mathbf{p}_l - \mathbf{t}) = \mathbf{R}_r\mathbf{R}_l^T(\mathbf{p}_l - \mathbf{t}_l + \mathbf{R}_l\mathbf{R}_r^T\mathbf{t}_r), \tag{4.42}$$

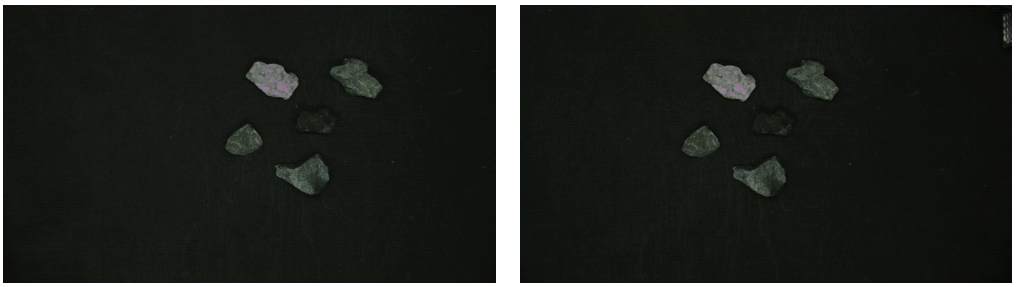
from which we can identify that

$$\mathbf{R} = \mathbf{R}_r\mathbf{R}_l^T \tag{4.43}$$

$$\mathbf{t} = \mathbf{t}_l - \mathbf{R}_l\mathbf{R}_r^T\mathbf{t}_r \tag{4.44}$$

describe the rigid transformation between the left and right camera frames in terms of extrinsic parameters obtained from camera calibration [22].

4.3 Rectification



(a) Camera 1 (left) (b) Camera 2 (right)

Figure 4.4: Raw image input from a stereo camera system. Images reproduced with permission from Antmicro Ltd[2].

[2] `https://antmicro.com/blog/2020/02/industrial-3d-camera-system/`, 2020 [Online; accessed 26-April-2021].

A view from a stereo camera system from different image planes from each camera is depicted in Fig 4.4. As visualized earlier in Fig. 4.3, epipolar lines are usually not collinear, complicating the search for matching points computationally. This is solved by projecting the images onto a common image plane, given by projective transformations $\mathbf{H}_l$ and $\mathbf{H}_r$. Additionally, this transformation is done with the constraint that epipolar lines must be parallel to the horizontal axis of the common image plane and corresponding points must have identical vertical coordinates. This simplifies the search for matching points to the horizontal axis.

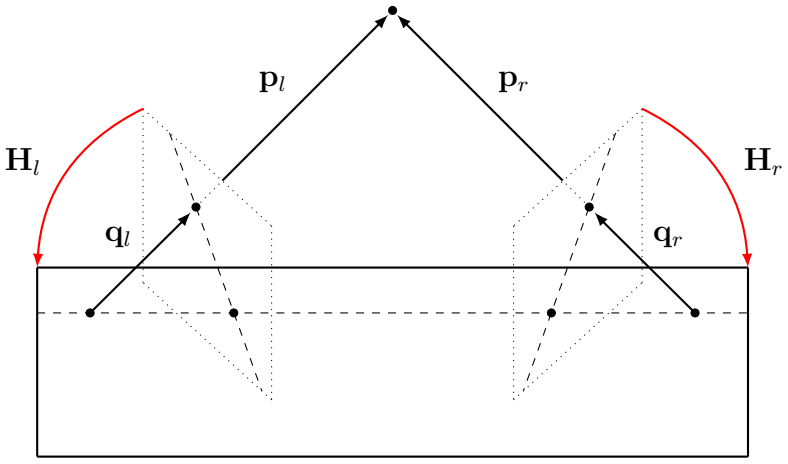


Figure 4.5: Image rectification procedure. Images are projected to a common plane where epipolar lines are parallel to the horizontal axis.

From Eq. (4.43) we know the relative rotation $\mathbf{R}$ and translation $\mathbf{t}$ between the cameras. Rotation matrix $\mathbf{R}$ can be transformed to an axis-angle representation describing the axis of rotation and its magnitude. Using this representation, we can split the relative rotation to two rotations, described by rotation matrices $\mathbf{R}_l$ and $\mathbf{R}_r$ which are obtained using Rodrigues' rotation formula from the axis-angle representation. In this case, each camera rotates a half rotation, such that the cameras are rotated into a coplanar alignment [23].

This does not guarantee that epipolar lines are parallel to the horizontal axis. A row alignment rotation can be constructed by the choice of a new coordinate system. For the first vector, we choose the direction of the epipole as

$$\mathbf{e_1} = \frac{\mathbf{t}}{\|\mathbf{t}\|}. \tag{4.45}$$

Furthermore, we require that $\mathbf{e_2}$ is orthogonal to the first vector. This can be obtained by computing and normalizing a cross product between the optical

axis direction vector and $\mathbf{e_1}$ as

$$\mathbf{e_2} = \frac{1}{\sqrt{T_x^2 + T_y^2}} \begin{bmatrix} -T_y & T_x & 0 \end{bmatrix}^T . \tag{4.46}$$

Finally, the third unit vector is obtained by the cross product

$$\mathbf{e_3} = \mathbf{e_1} \times \mathbf{e_2}. \tag{4.47}$$

Row alignment is then achieved by the rotation matrix

$$\mathbf{R}_{\text{rect}} = \begin{bmatrix} \mathbf{e_1}^T \\ \mathbf{e_2}^T \\ \mathbf{e_3}^T \end{bmatrix}, \tag{4.48}$$

which is applied to both camera views after rotating them to coplanar alignment [22].

Homographies $\mathbf{H}_l$ and $\mathbf{H}_r$ are then constructed by

$$\mathbf{H}_l = \mathbf{K}\mathbf{R}_{\text{rect}}\mathbf{R}_l\mathbf{K}_l^{-1} \tag{4.49}$$

and

$$\mathbf{H}_l = \mathbf{K}\mathbf{R}_{\text{rect}}\mathbf{R}_r\mathbf{K}_r^{-1}, \tag{4.50}$$

where $\mathbf{K}$ represents a common intrinsic matrix for both cameras, which can be chosen arbitrarily [24]. Figure 4.6 depicts the result of the image rectification procedure, with resulting left and right images overlaid on top of each other.

Figure 4.6: Rectified and overlaid image. Image reproduced with permission from Antmicro Ltd.

4.4 Disparity and depth

With images being projected to a common plane with a common focal depth, depth computation becomes trivial. The process follows a simple triangulation method using two parameters obtained earlier: baseline $B = o_r - o_l$ and focal depth f. Focal depth is the effective focal depth in pixels (recall definitions of f_x and f_y in K, physical focal depth was scaled to pixels), and baseline is in physical units. Let us assume we have identified positions u_r and u_l from the rectified image that correspond to the same point $\mathbf{p}$

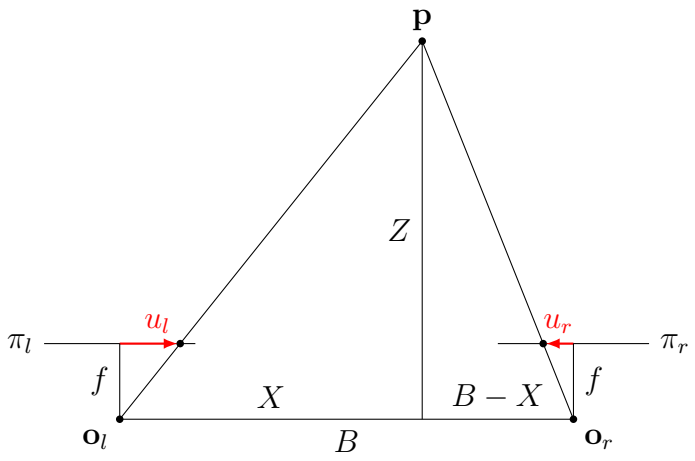


Figure 4.7: Triangulation geometry containing two pairs of similar triangles.

Figure 4.7 displays two pairs of similar triangles, from which we can identify that

$$\frac{Z}{f} = \frac{X}{u_l} \quad \text{and} \quad \frac{Z}{f} = \frac{B - X}{u_r} \tag{4.51}$$

Solving for *disparity* d gives

$$d = u_r + u_l = \frac{f(B - X)}{Z} + \frac{fX}{Z} = \frac{fB}{Z}, \tag{4.52}$$

and finally, solving for depth Z we have

$$Z = \frac{fB}{d}. \tag{4.53}$$

So far we have only discussed single points of interest and we have assumed we have identified their locations in both images. This is not a fair assumption to make in practice, and the problem of correspondence has to be

addressed here. The OpenCV library [25] implements a semi-global matching based algorithm [26], which provides a dense disparity map from a rectified stereo image pair. This implementation is outside the scope of the thesis, but it is suitable for real-time applications.

Figure 4.8: Dense disparity image provided by semi-global matching. Image reproduced with permission from Antmicro Ltd.

As it is evident from Fig. 4.8, a distinct background plane is visible. Disparity is in terms of pixels, and depth is relative to the camera baseline. The objective is to measure sample thickness. A random sample consensus (RANSAC) [27] based algorithm was implemented that, in the presence of noise, fits a robust estimate for a background plane of the form

$$Ax + By + Cz + D = 0, \tag{4.54}$$

to the depth image. Sample thickness was then simply computed by subtracting depth map from the background plane.

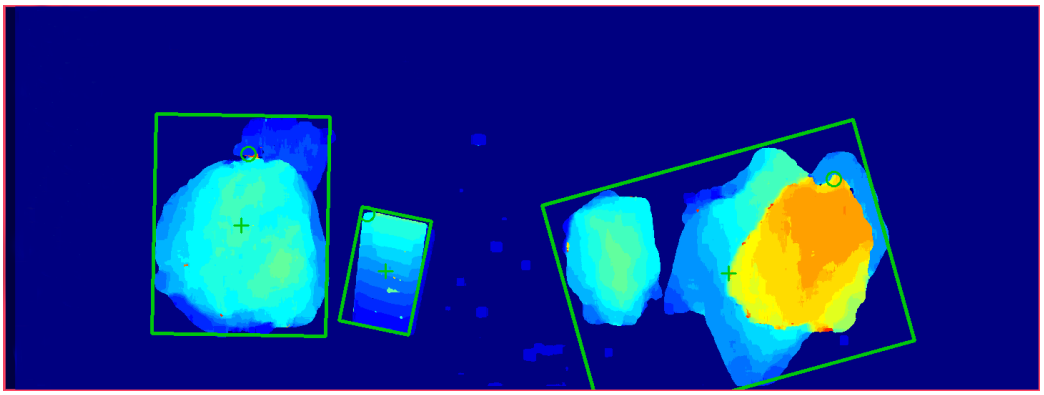


Figure 4.9: Background plane corrected depth map, with additional bounding boxes demonstrating object detection. Image reproduced with permission from Antmicro Ltd.

Chapter 5

Data fusion

5.1 Image registration

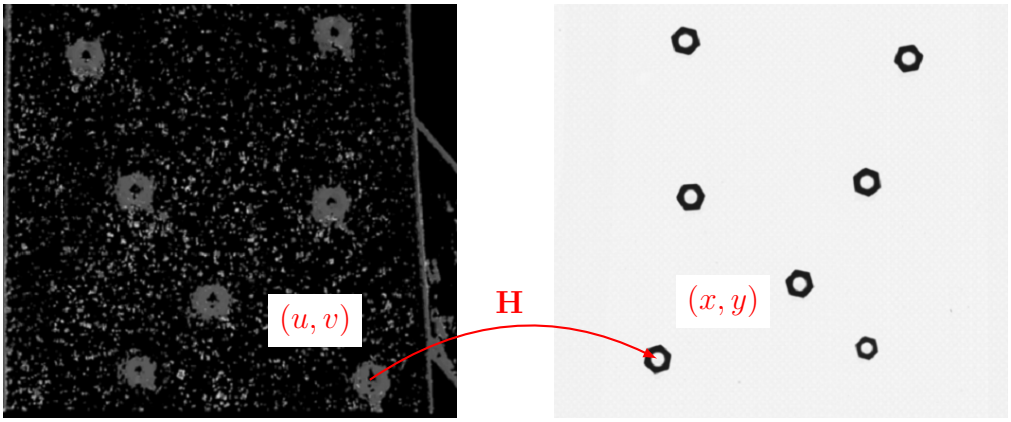
Images produced by Medipix3 and stereo camera are formed on their respective image planes. An image registration procedure is needed in order to align the data into a common coordinate plane for data fusion.

5.1.1 Projective transformation

Both cameras are assumed to follow a pinhole camera model. For Medipix3, this follows immediately as no imaging optics are used. Stereo camera image output is free of distortion due to camera calibration. Due to pinhole camera model, a homography **H** relates the projective transformation between the image planes formed by Medipix3 and stereo camera. A point (u, v) in the disparity image plane can be transformed to X-ray plane (x, y) by the projective transformation **H** as

$$\begin{bmatrix} \lambda x \\ \lambda y \\ \lambda \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix},$$

where h_{ij} denote components of the homography matrix **H** and λ is the scale factor.



a) Disparity image b) X-ray image

Figure 5.1: Example of a point correspondence using circular control points as features. Result is a coordinate pair of (u, v) and (x, y) representing the same point in physical space.

Following the homography estimation method described in Section 4.1.1 and constraining homography scale to $h_{33} = 1$, we have

$$\begin{bmatrix} \vdots \\ x_i \\ y_i \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots & & \vdots & & \vdots & & \vdots & \\ u_i & v_i & 1 & 0 & 0 & 0 & -x_i u_i & -x_i v_i \\ 0 & 0 & 0 & u_i & v_i & 1 & -y_i u_i & -y_i v_i \\ \vdots & & \vdots & & \vdots & & \vdots & \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} \tag{5.1}$$

or

$$\mathbf{a} = \mathbf{M}\mathbf{h} \tag{5.2}$$

for each point correspondence between the images, where $\mathbf{M}$ is a $2n \times 8$ matrix. The process of obtaining corresponding points from both image planes is demonstrated in Fig. 5.1. Again, assuming we have at least $n \geq 4$ corresponding points from both image planes, a linear least squares solution gives us

$$\mathbf{h} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{a}, \tag{5.3}$$

where vector $\mathbf{h}$ contains the matrix elements for constructing homography $\mathbf{H}$.

5.1.2 Interpolation

Since Medipix3 images objects at a higher image resolution when compared to the stereo camera, we wish to compute a result that retains this result but includes information from both cameras. Transformation **H** provides the necessary information for transforming each individual pixel in the disparity image to the X-ray image plane. However, due to image resolution mismatch, directly mapping disparity image pixels to X-ray plane does not provide an unambiguous result on a pixel-to-pixel basis.

This can be solved by inverting the homography **H** and iterating over every pixel coordinate from X-ray image plane and projecting to disparity image plane, and using bilinear interpolation to assign a value for each X-ray pixel. Intensity estimation using bilinear interpolation is visualized in Fig. 5.2.

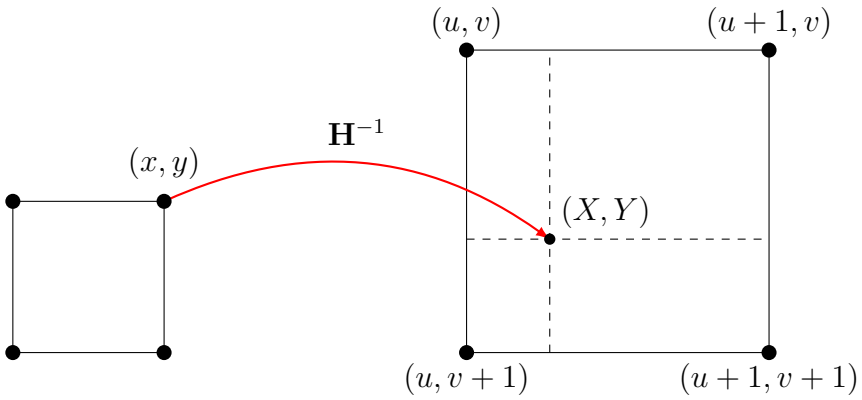


Figure 5.2: Intensity estimation at (X, Y) using bilinear interpolation with intensities at (u, v), $(u+1, v)$, $(u, v+1)$, and $(u+1, v+1)$. Coordinates such as (u, v) represent pixels in the original disparity image. Coordinate (X, Y) represents a transformed pixel from X-ray image using transformation $\mathbf{H}^{-1}$.

We have

$$\begin{aligned} I(x, y) = & W_{u,v} I(u, v) \\ & + W_{u+1,v} I(u+1, v) \\ & + W_{u,v+1} I(u, v+1) \\ & + W_{u+1,v+1} I(u+1, v+1), \end{aligned} \tag{5.4}$$

where

$$W_{u,v} = (u + 1 - X)(v + 1 - Y), \tag{5.5}$$
$$W_{u+1,v} = (X - u)(v + 1 - Y), \tag{5.6}$$
$$W_{u,v+1} = (u + 1 - X)(Y - v), \tag{5.7}$$
$$W_{u+1,v+1} = (X - u)(Y - v) \tag{5.8}$$

and I represents the pixel intensity at specified position [28]. Resulting composite image is shown in Fig. 5.3, where transformed and interpolated disparity image is overlaid on top of the X-ray image.

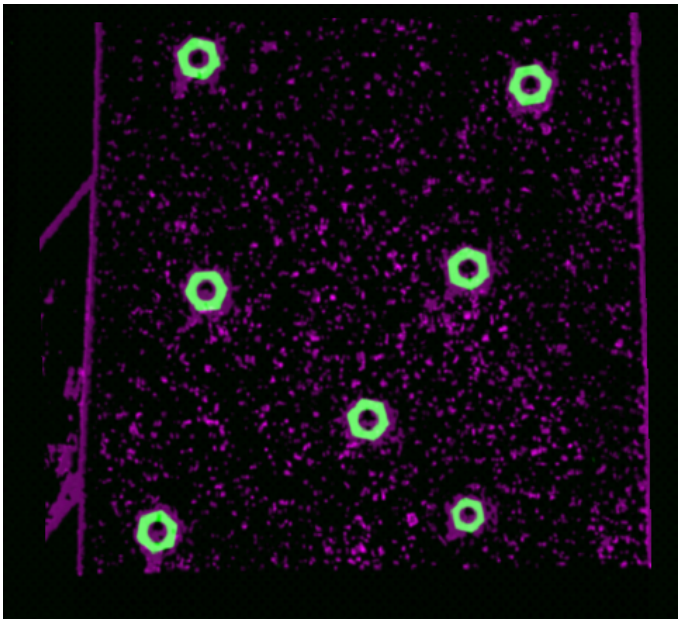


Figure 5.3: Composite image of X-ray and disparity image registration.

Chapter 6

Algorithms

This chapter outlines two algorithms and a method for evaluating their performance. An algorithm using a combination of single channel mode and stereo vision is compared to an algorithm using dual channel mode. Both algorithms produce a single numerical value for every pixel.

6.1 Single channel and thickness

Here, only a single channel is used on Medipix3 at 12 bits of counter depth for each pixel with energy threshold E_0. If we measure a flat-field image with no sample present, the measured intensity is

$$I_0 = \int_{E_0}^{E_{\max}} I_s(E)D(E)\,dE, \tag{6.1}$$

and subsequently measure a sample with intensity

$$I = \int_{E_0}^{E_{\max}} I_s(E)D(E)e^{-\sum_{i=1}^{N}\mu_i(E)\Delta x_i}\,dE. \tag{6.2}$$

We may attempt to approximate the relation between I and I_0 with the exponential law

$$I = I_0 e^{-\mu x}, \tag{6.3}$$

where x is obtained from thickness image using stereo camera. Solving for μ, we have

$$\mu = -\frac{1}{x}\ln\big(\frac{I}{I_0}\big), \tag{6.4}$$

where μ represents a numerical estimate for an effective attenuation coefficient, and an estimate for the material.

6.2 Dual channel

With two discriminators, we can measure two portions of the spectrum, defined by energy thresholds E_0 and E_1. For the first channel, we measure an intensity

$$I_0 = \int_{E_0}^{E_{\max}} I_s(E)D(E)e^{-\sum_{i=1}^{N}\mu_i(E)\Delta x_i}\,dE, \tag{6.5}$$

and for the second channel

$$I_1 = \int_{E_1}^{E_{\max}} I_s(E)D(E)e^{-\sum_{i=1}^{N}\mu_i(E)\Delta x_i}\,dE. \tag{6.6}$$

As denoted in Section 3.3.2, the intensitites share a common portion of the spectrum above the higher threshold. We obtain independent measurements by defining channel responses as

$$c_0 \equiv I_0 - I_1 \tag{6.7}$$

and

$$c_1 \equiv I_1. \tag{6.8}$$

Assuming an ideal detector[1], we can simulate the macroscopic behaviour of sample attenuation using the spectrum obtained from SPEKTR 3.0 toolbox [4], together with elemental attenuation coefficients obtained from XCOM database [12]. For copper and aluminium, the parametric plot of intensity as a function of sample thickness is visualised in Fig. 6.1.

[1]With ideal absorption efficiency and no charge sharing or alternative degrading effects on imaging

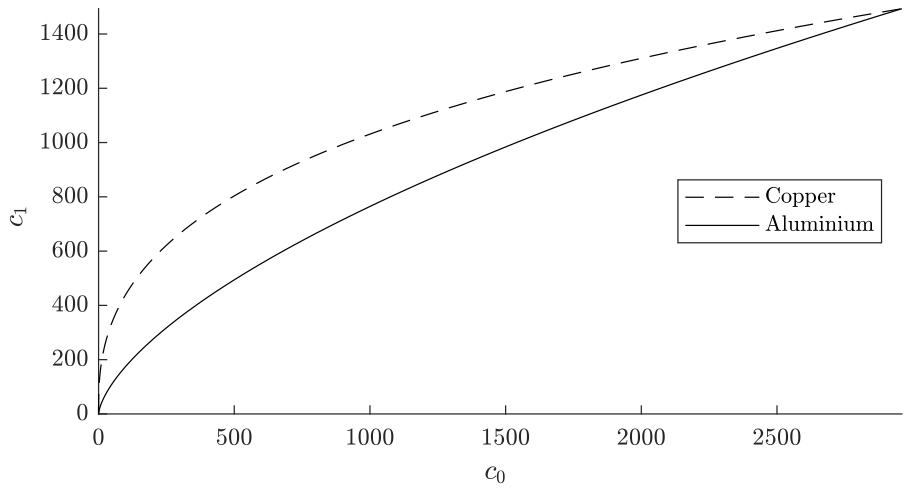


Figure 6.1: Parametric plot of channel responses as a function of sample thickness for copper and aluminium samples.

Initially, the channel responses c_0 and c_1 start from a common point, which corresponds to the intensity with no sample present. Due to differences in spectral attenuation, the parametric curves quickly deviate from one another, until they meet again at the point where intensity has been reduced to zero for both channels. The form of the parametric curve is dependent on threshold selection.

We now employ signal-to-equivalent thickness calibration discussed in Section 3.5, using aluminium as reference material. Calibration is done for both channels c_0 and c_1 by measuring aluminium samples with varying thicknesses. Following the calibration, we may transform the measured intensities to equivalent thickness as

$$t_0 \equiv t_0(c_0) = t_0(y_0 - y_1) \tag{6.9}$$

and

$$t_1 \equiv t_1(c_1) = t_1(y_1). \tag{6.10}$$

This effectively linearizes the response relative to the calibration material. This is shown in Fig. 6.2, where the calibration material is diagonal and other materials deviate from the diagonal line.

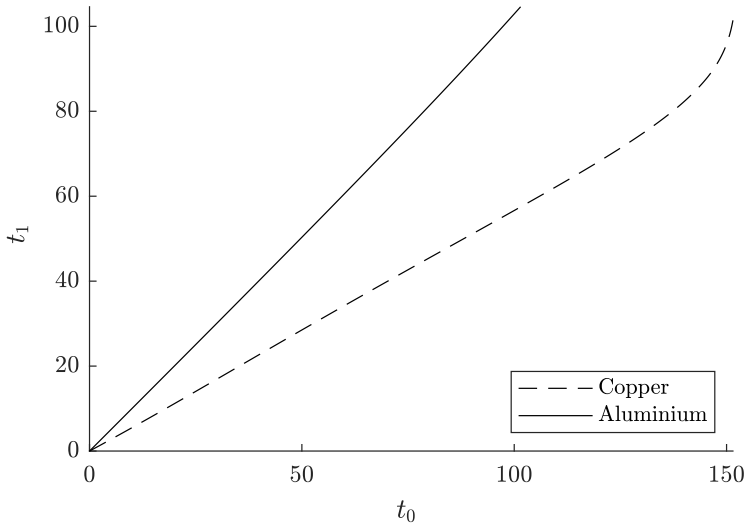


Figure 6.2: Result of signal-to-equivalent thickness calibration. Reference material (aluminium) is diagonal.

Finally, transforming the data by

$$t = \frac{1}{2}(t_0 + t_1) \tag{6.11}$$

and

$$m = t_0 - t_1, \tag{6.12}$$

yields $m = 0$ for reference material as $t_0 = t_1$. Deviation from $m = 0$ can therefore be interpreted as different material.

Let us briefly discuss the qualitative behaviour of the material estimate m as a function of atomic number Z, assuming aluminium (Z_{Al}=13) as calibration material and an ideal detector. As Z increases, lower energy photons are more effectively attenuated relative to the calibration material. This implies that c_0 reduces at a faster rate, resulting in a higher thickness estimate t_0. Thickness estimate t_1 is less affected by this, and as a result m increases.

Figure 2.5 depicts the mass attenuation coefficients as a function of Z and photon energy E. Due to photoelectric effect, there are asymmetric discontinuities, describing different characteristic edges. Above the characteristic edge, photons are emitted as characteristic photons. For aluminium, characteristic K edge is at $E_K = 1.5$ keV. As Z increases, E_K increases. At some point, E_K approaches the lower threshold E_0, converting photons from the lower channel c_0 to characteristic photons. These photons were originally

above the threshold, but conversion to characteristic photons results in a rejection by the discriminator. This results in a decrease in c_0 and an increase in t_0, increasing the material estimate even further. As E_K passes E_0, characteristic photons are accepted by the lower energy discriminator, resulting in a rapid increase in c_0. This reduces lower channel thickness estimate, and material estimate m begins to rapidly decrease. Once E_K approaches the upper threshold E_1, photons are converted from c_1 to c_0 again due to photoelectric effect. Counts in c_0 increase rapidly, and thickness estimate t_0 decreases rapidly, vice versa for the upper channel. This decreases the material estimate even more. As E_K passes E_1, characteristic photons no longer have as significant role in m, and the original behaviour resumes with m increasing as Z increases.

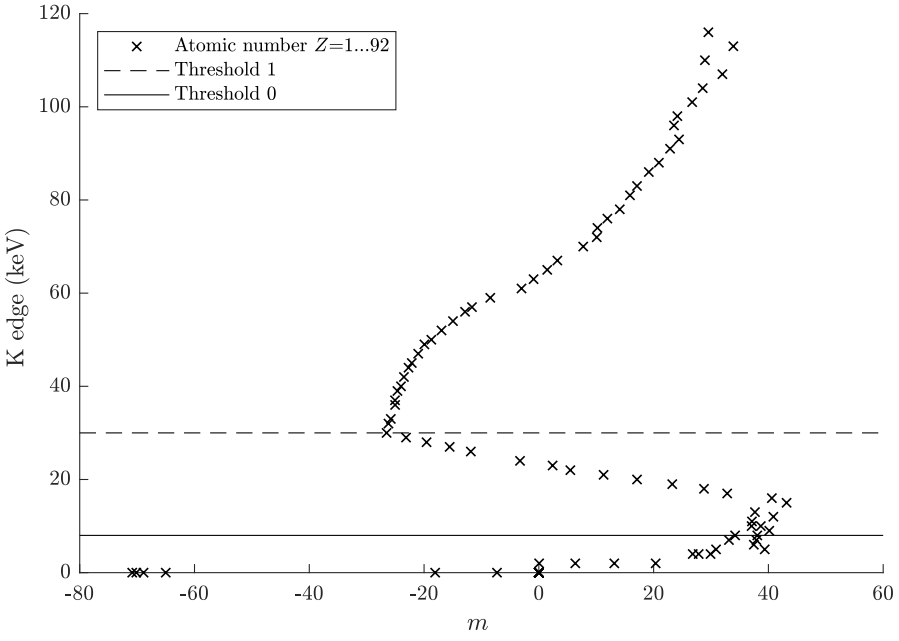


Figure 6.3: Simulated behaviour of material estimate as a function of atomic number K edge. Average material thickness t is constant. Thresholds are E_0 = 10 keV and E_1 = 30 keV.

As it is evident from Fig. 6.3, material estimation is not bijective. Threshold selection defines degeneracies between materials. Additionally, the issues of charge sharing degrade energy resolution. For real-time measurements, dual channel approach requires more data to be processed compared to single channel approach. This increases the readout time and decreases the

available integration time for the detector, reducing intensities measured and introducing higher photon noise. As a result, single channel approach can have a higher throughput in terms of samples processed over time, considering the effect of photon noise.

6.3 Evaluation

Algorithms described earlier both operate on a pixel basis. However, a method to evaluate their performance is still needed. For the purposes of real time material discrimination, samples were obtained from various European mines in collaboration for determining the performance. The samples were individually measured using an X-ray tomography based GeoCore X10 drill core scanner, developed by Orexplore AB[2]. Drill core scanner provided a non-destructive method for obtaining reference data for each sample, on an elemental basis, where mass percentage estimates for each element were provided.

Evidently, from the perspective of retaining as much target material as possible, the optimal solution is to retain all samples when performing sample discrimination. The balancing force is the economical perspective, where a sample might contain too little target material to be worth processing. This is a function of the ratio of ore material relative to the waste material that needs to be processed. In mathematical terms, this forms a threshold t_a that is related to the number of pixels classified as ore relative to the area of the sample in pixels. Both algorithms described in this chapter aim to provide a single numerical value for each pixel, which can be also thresholded to an ore or waste classification. This forms a set of two thresholds (t_a, t_p), where t_a is derived from an economical perspective, and t_p is derived from a physical perspective. Assuming we have a set of samples that somewhat accurately describes the population of samples, we can form an optimized estimate for t_p for both algorithms. Therefore, only t_a is left to be optimized, determined entirely by economical benefit.

A common approach to describe the diagnostic ability of a binary classifier system is to compute a receiver operating characteristic curve (ROC). In principle, a similar curve is introduced here. With the reference data, we can sort the samples based on ore content and compute a theoretical upper bound for the performance. We compute a ratio of ore mass with respect to the total mass of each sample, and iterate the threshold t_p over a range where no sample is classified as ore to the limit of every sample being classified as

[2] `https://orexplore.com/`

ore. This results in a plot of ore mass kept as a function of total mass kept. A set of curves is produced. Threshold t_p is chosen by selecting the corresponding curve that maximizes the area-under-curve and this is assumed to be an optimized value for threshold t_p.

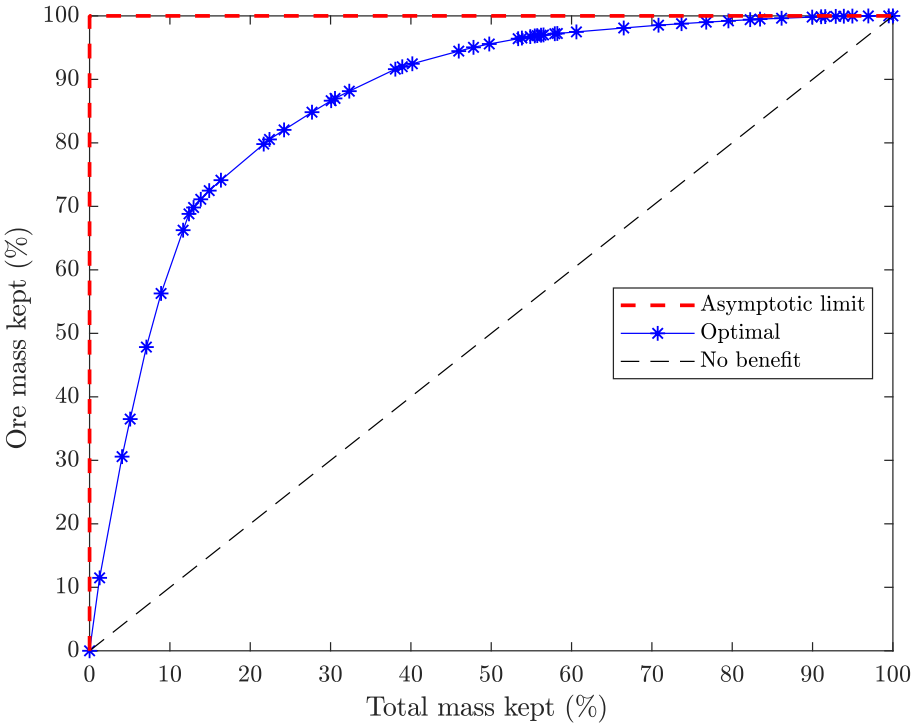


Figure 6.4: Optimal curve for algorithm performance, obtained from maximizing area-under-curve using reference data from GeoCore X10.

Figure 6.4 depicts an optimal curve corresponding to threshold t_p. The asymptotic limit demonstrates a case where samples are clearly divided into ore and waste, and ore mass contributes little to the total mass. In this case, a great deal of waste can be evidently removed from the ore processing at an early stage. On the contrary, a diagonal line demonstrates the opposite scenario, where samples are so uniformly distributed in ore content that sorting cannot introduce any benefit. An algorithm operating on the same set of samples should produce a curve below the optimal curve. The optimal curve serves as an ideal case for algorithm performance, since the limiting factor is ore content distribution within samples.

Chapter 7

Instrumentation, materials and measurements

The purpose of this chapter is to give the reader a more concrete overview of the instrumentation, samples and the conducted measurements, before committing to the results in the next chapter.

7.1 Devices and hardware

A stereo vision camera system for the purpose of real-time sample thickness measurements was developed by Antmicro Ltd during the project. The device is shown in Figure 7.1 and the camera details are outlined in Table 7.1. Real-time operation is achieved with GPGPU[1] acceleration operating with the open source algorithms mentioned in Section 4.4. The camera is fitted with two 16 mm focal length, electronically controlled liquid lenses that enable a focusing range of 8 cm.

[1]General-Purpose Graphics Processing Unit

Figure 7.1: Industrial 3-D stereo vision camera system developed by Antmicro Ltd.

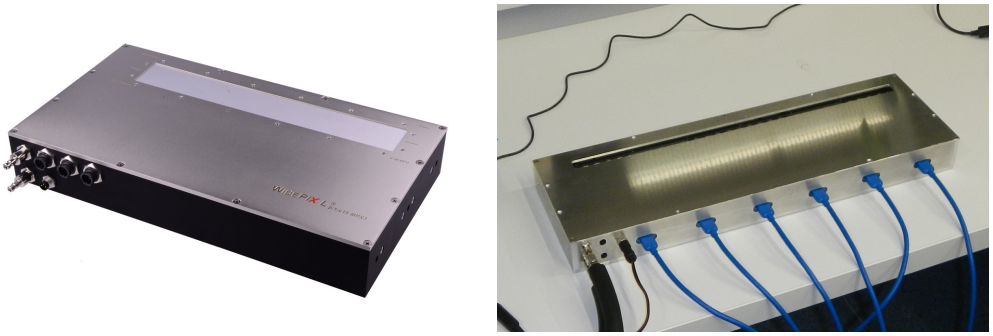
The distance from belt to camera is 80 cm, and the viewing area is 50x25 cm. From the parameters obtained from camera calibration, the focal depth in pixels is roughly 2909 pixels. With sample thicknesses up to 5 cm, we have a range for distances from 75 cm to 80 cm. From Eq. (4.53), we obtain a range for disparities from 182 to 194 with a bit depth of 8 per pixel. This yields a depth resolution of roughly 4 mm/disparity level. For the purpose of improving depth resolution, a sub-pixel estimation algorithm was implemented by manipulating input images (horizontal stretching with linear interpolation) and subsequently scaling the disparity map. In addition, the device supports external hardware triggering for synchronizing the device operation.

Sensor	Resolution (px)	Baseline (cm)	Height (cm)	Frame rate	Interface
CMV2000	2048 x 1088	5	80	20	Ethernet

Table 7.1: Stereo vision camera details.

As prefaced in Chapter 1, the demanding requirements of the application introduced a need for high-speed electronics readout for the Medipix3 array. Developed Widepix[2] devices are depicted in Fig. 7.2 and the device details are outlined in Table 7.2.

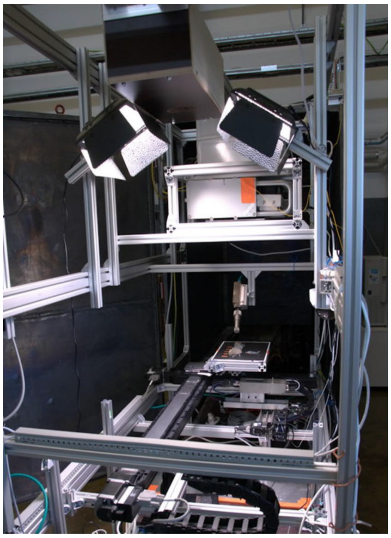
[2]Medipix3 array



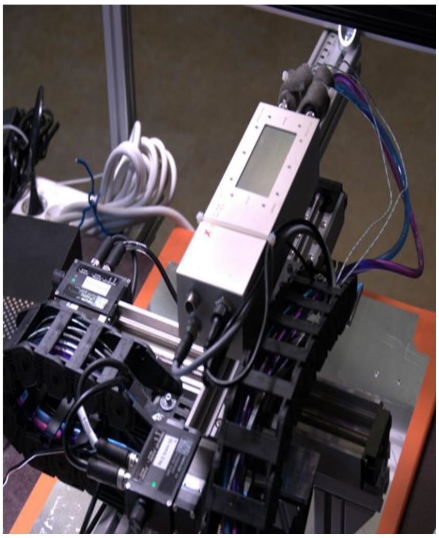
(a) Widepix L 1x15. (b) Widepix F 1x30.

Figure 7.2: Images of 210 mm (left) and 420 mm (right) Widepix devices. New high-speed readout architecture is included in the 420 mm device.

Initial tasks related to research and development of the real-time data fusion were conducted at the VTT X-ray laboratory in Oulu, Finland. A radiation shielded X-ray cabinet containing various remotely controlled linear actuators, a stereo vision camera system and a Widepix 1x5 is shown in Fig. 7.3. The linear actuators are capable of simulating conveyor belt behaviour up to 1.5 m/s, but the practical limitation is due to old readout electronics architecture of the Widepix 1x5 device. The sample tray is attached to a linear actuator and can be moved at selected speed between the Widepix and stereo vision camera for imaging and data fusion purposes. A real-time demonstration using this configuration is illustrated in Section 8.4.



(a) Overview.



(b) Widepix 1x5.

Figure 7.3: Interior of an X-ray cabinet at the VTT laboratory. Linear actuators enable sample movement in order to simulate conveyor belt behaviour. Stereo vision camera system is installed with additional radiation shielding.

Camera	Sensitive area (cm)	Resolution (px)	Frame rate	Interface
Widepix 1x5	7 x 1.4	1280 x 256	50	USB2
Widepix L 1x15	21 x 1.4	3840 x 256	50	USB2
Widepix F 1x30	42 x 1.4	7680 x 256	450	USB3

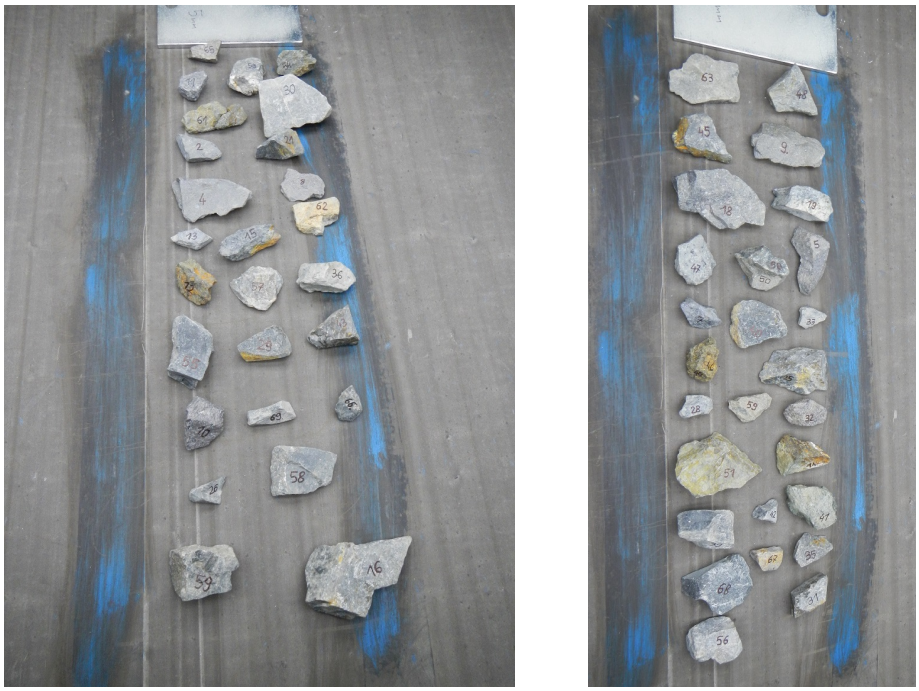
Table 7.2: Widepix device details.

7.2 Samples

A set of samples were obtained from three different mines: Lovisagruvan, Hellas Gold and Asarel Medet. Lovisagruvan is a small Swedish mine producing high grade zinc-lead-silver ore. In total, 53 samples were obtained from Lovisagruvan. Based on the reference data acquired using GeoCore X10, the samples contain mostly lead and zinc as ore elements. The metal content of the heavy elements between samples varied significantly, and the ranges are presented in Table 7.3. Samples for Lovisagruvan are shown in Figure 7.4.

Element	Minimum (%)	Maximum (%)	Average (%)
Zn	0.00	22.11	2.67
Pb	0.00	12.32	1.05
Fe	0.67	17.10	4.62

Table 7.3: Mass percentages of dominant heavy elements in the provided set of samples for Lovisagruvan.



(a) Batch 1 (b) Batch 2

Figure 7.4: Lovisagruvan samples.

The second mine, Hellas Gold, consists of two mining operations: a gold-silver-lead-zinc mine and a silver-lead-zinc mine, located in northern Greece. Similar to Lovisagruvan mine, the samples obtained contained mostly zinc and lead as ore elements. In total, 56 samples were obtained with reference data, with Table 7.4 describing the ranges of relevant heavy elements. Samples for Hellas Gold are displayed in Figure 7.5.

Element	Minimum (%)	Maximum (%)	Average (%)
Zn	0.03	23.97	2.61
Pb	0.00	16.74	0.81
Fe	1.41	30.34	10.8

Table 7.4: Mass percentages of dominant heavy elements in the provided set of samples for Hellas Gold.

(a) Batch 1 (b) Batch 2

Figure 7.5: Hellas Gold samples.

Asarel Medet is a Bulgarian mine, mining and processing mainly copper ore. Significant challenges in terms of ore sorting are due to relatively high iron content, and its similarity to copper in atomic number and therefore attenuation characteristics. In total, 55 samples were obtained and their copper and iron ranges are presented in Table 7.5. Samples for Asarel Medet are displayed in Figure 7.6.

Element	Minimum (%)	Maximum (%)	Average (%)
Cu	0.00	1.32	0.14
Fe	1.32	7.37	3.58

Table 7.5: Mass percentages of dominant heavy elements in the provided set of samples for Asarel Medet.

Figure 7.6: Asarel Medet samples.

7.3 Measurements

Measurements were divided to two separate campaigns using a sorter research prototype at a Comex Group[3] laboratory in Poland using a research sorter prototype, which is shown in Figure 7.7.

(a) Overview. (b) Stereo vision camera.

Figure 7.7: Overview of the research sorter prototype at Comex Group laboratory in Poland, where the measurement campaigns were conducted. The right image features the stereo vision system installed with additional illumination.

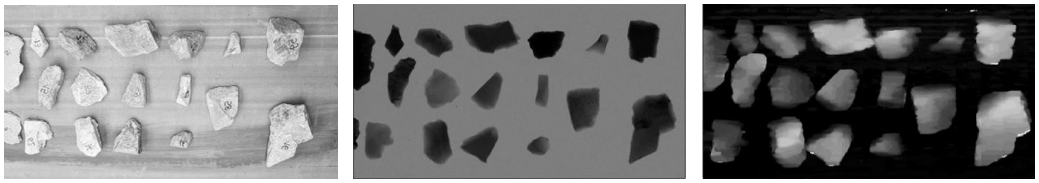
The first campaign employed the 210 mm wide Medipix3 array using single channel mode, augmented with a stereo vision camera system. A subsequent secondary campaign demonstrated the high-speed readout electronics for a full 420 mm wide array, imaging the same set of samples with dual channel mode at much higher speeds. Parameters, such as integration time t, sample movement speed v, energy thresholds E_0 and E_1 are laid out in Table 7.6. X-ray tube voltages were 160 kVp for both campaigns, with 2.0 and 2.5 mA tube currents respectively. Measurement parameter choices were strictly driven by the requirements of real-time operation, with the goal of minimizing X-ray photon noise. Due to charge sharing effects, measured incident spectrum is skewed towards lower energies and the use of higher energy thresholds reduces the measured intensity. Thresholds were not optimized in any way other than to maximize and equalize channel counts between channels.

[3] `http://www.comex-group.com`

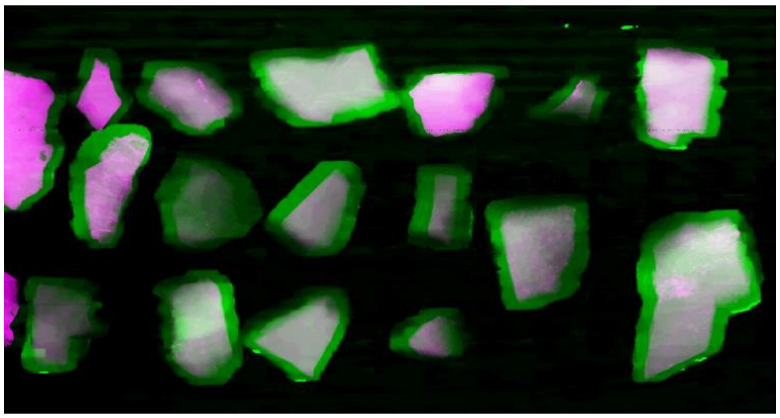
Channels	Width (mm)	E_0 (keV)	E_1 (keV)	t (ms)	v (m/s)	I (mA)
Single	210	15	-	50	0.24	2.0
Dual	420	8	20	1	1.50	2.5

Table 7.6: Measurement parameters for the two campaigns. High speed and low photon counts restrict the choice of energy thresholds. Thresholds were chosen with the intent to equalize and maximize photon counts between channels.

As visualized in Figure 5.1, image registration between cameras was implemented using circular features obtained from metal nuts, where the circles and the circle centers are the points of interest in both image planes. However, the first campaign did not incorporate synchronized operation, and the disparity images were manually processed to mimic a result from synchronization. After image registration, only a single free parameter is left, which corresponds to the offset between images along the belt movement direction. This offset was manually chosen for each disparity image and the result is shown in Figure 7.8. In addition, the sub-pixel estimation was configured to improve depth resolution threefold.



(a) Image of samples. (b) X-ray image. (c) Disparity image.



(d) Composite image.

Figure 7.8: Offline synchronization results for data fusion in the first measurement campaign.

The dual channel algorithm required the measurement of reference material, and aluminium was chosen as it is an adequate representation of waste material. Figure 7.9 depicts the chosen calibration thickness steps, and the result of the signal-to-equivalent thickness calibration for both channels, as described in Section 3.5.

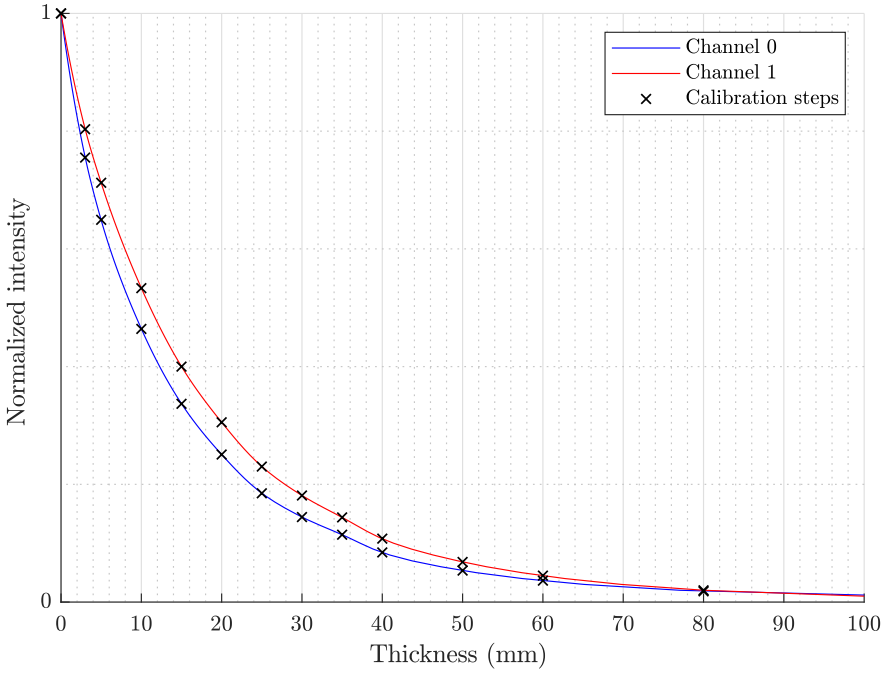


Figure 7.9: Reference material with various thicknesses for signal-to-equivalent thickness calibration. Aluminium was used as the reference material up to 80 mm thickness.

Chapter 8

Results

An upper bound for the performance is obtained from the optimal results provided by the drill core scanner. A baseline algorithm was obtained from simply employing the single channel approach without thickness compensation. Thickness based methods were divided to two approaches: first one applies a pixel-wise thickness compensation, and the second one simply computes a maximum thickness for each sample, which is applied pixel-wise. Additionally, a qualitative comparison to material estimation is visualized using the baseline algorithm.

Samples from the three individual mines were analyzed using both algorithms. Table 8.1 depicts the mines and their respective ore elements that are extracted. Other heavy elements, such as iron, can contribute to decrease in sorting performance due to similar atomic number to that of copper and zinc.

Mine	Ore elements	Other heavy elements
Lovisagruvan	Zn, Pb	Fe
Hellas Gold	Zn, Pb	Fe
Asarel Medet	Cu	Fe

Table 8.1: Classifications of ore in analysis for each specific mine. Contributions from ore mass for each ore element are summed.

8.1 Lovisagruvan

Figure 8.1 displays the results for the thickness-based algorithms. A clear improvement in performance is obtained by augmentation of thickness infor-

mation. The optimal curve signals that a good amount of waste mass can be rejected without a significant loss in ore mass. This is dictated by the choice of threshold t_a, which defines a set of points along the curve.

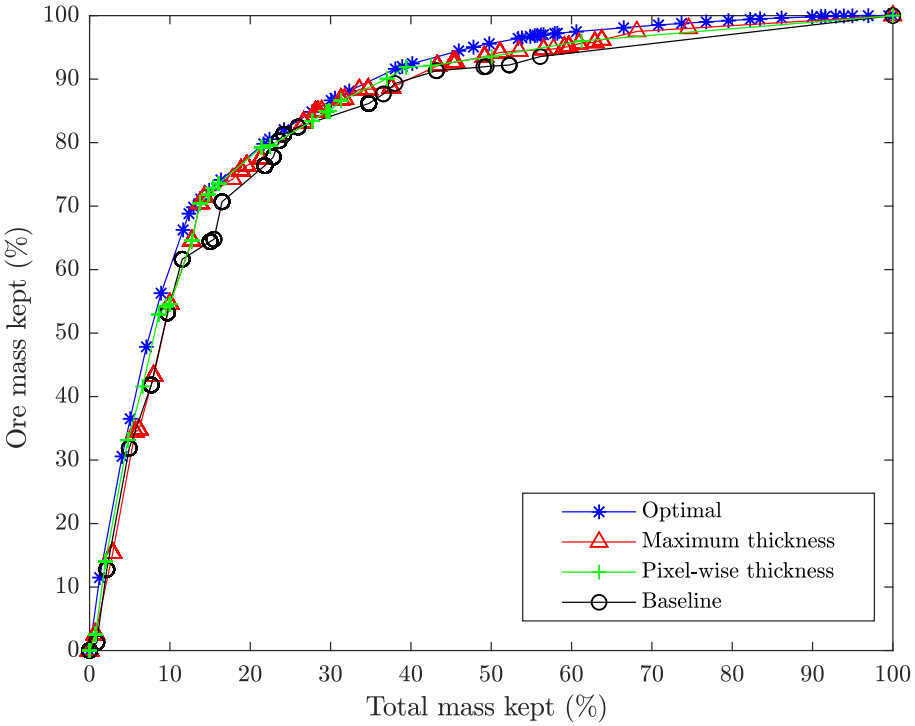


Figure 8.1: Lovisagruvan, first campaign.

Correspondingly, results of the dual channel approach from the subsequent campaign are depicted in Figure 8.2. Acquired performance is yet again arguably better than the baseline, if one considers that the choice of t_a is driven by the goal of high ore mass kept percentage.

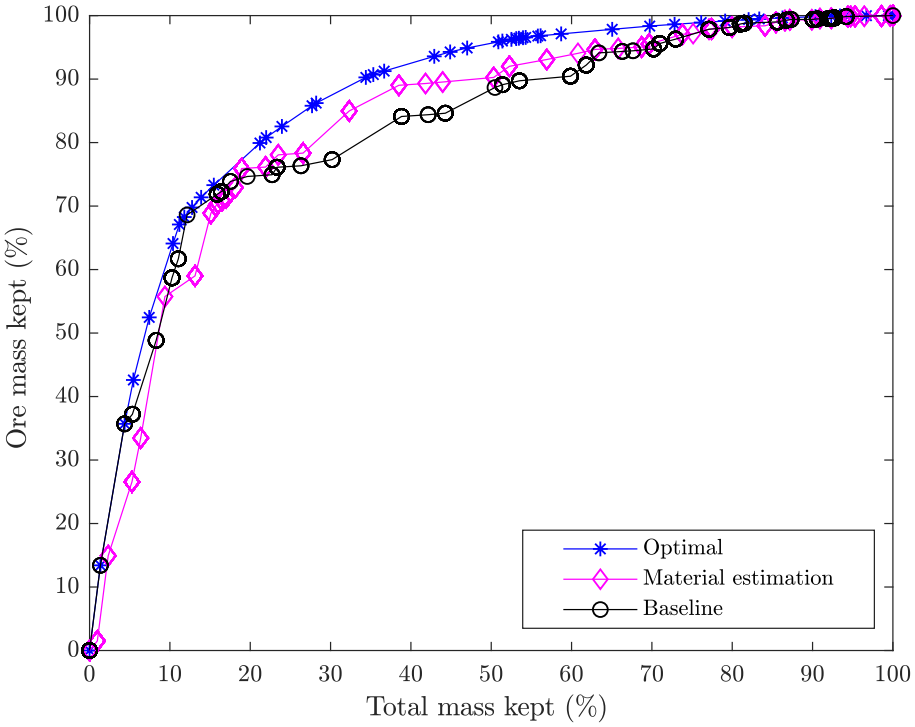


Figure 8.2: Lovisagruvan, second campaign.

8.2 Hellas Gold

For Hellas Gold, Figure 8.3 depicts the results from the first campaign. A clear improvement over baseline is achieved over the relevant range for threshold t_a, corresponding to high ore mass yield.

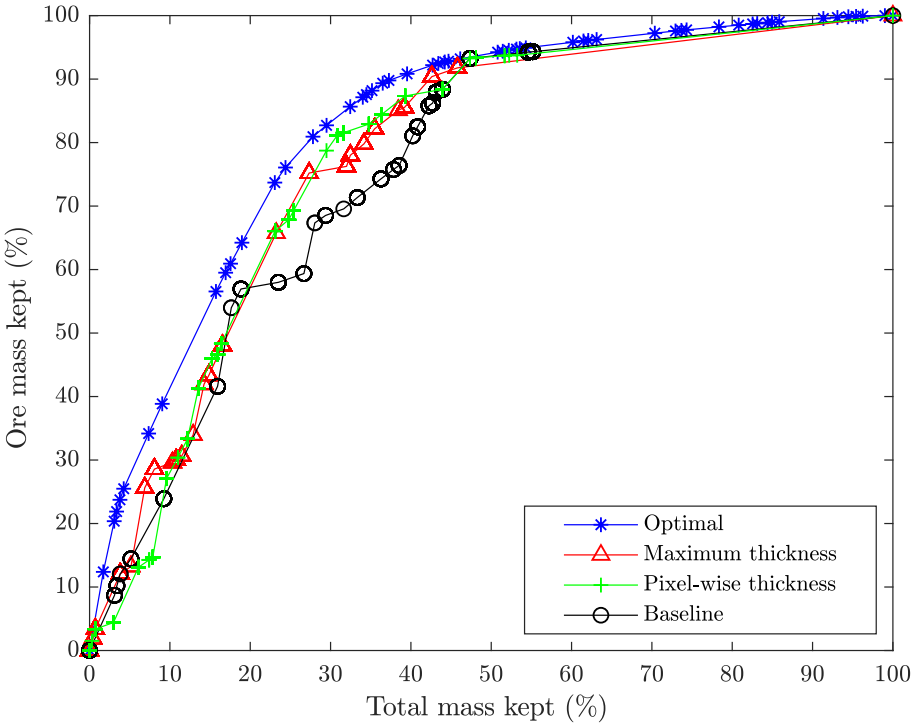


Figure 8.3: Hellas Gold, first campaign.

In contrast, dual channel approach is depicted in Figure 8.4, signalling difficulty in producing any additional information over the samples when compared to baseline.

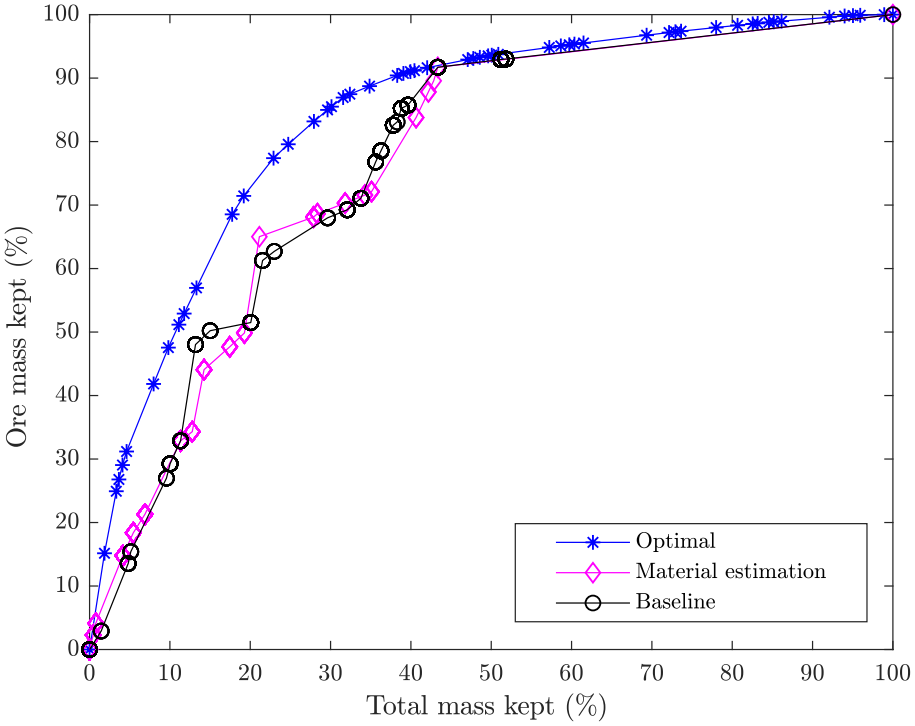


Figure 8.4: Hellas Gold, second campaign.

8.3 Asarel Medet

For Asarel Medet, the optimal curve signals that ore and waste distribution in samples is more uniform, and sorting cannot introduce a significant reduction in waste without significantly reducing ore mass simultaneously. As Figure 8.5 depicts, an improvement over the baseline is achieved, but with room for improvement.

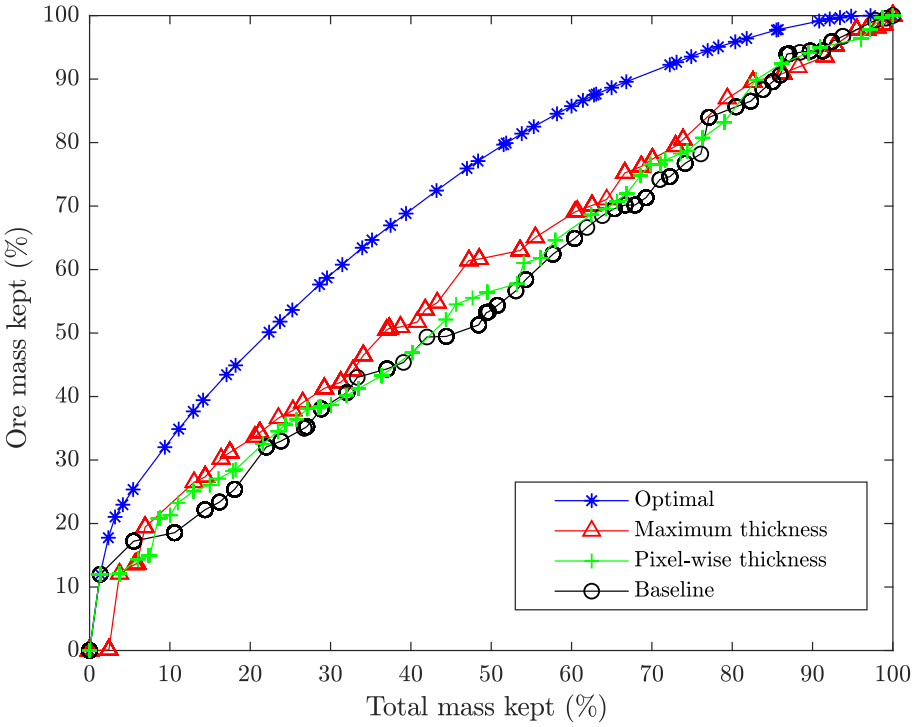


Figure 8.5: Asarel Medet, first campaign.

Similar to Hellas Gold, material estimation for Asarel Medet fails to produce any additional information over the baseline, as visualized in Figure 8.6.

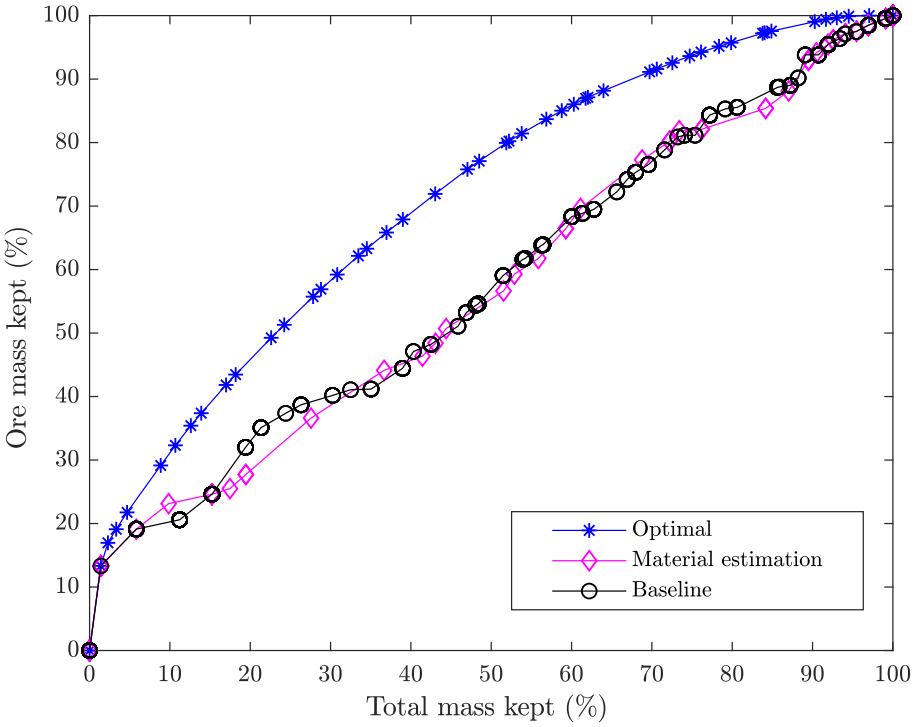


Figure 8.6: Asarel Medet, second campaign.

8.4 Online demonstration

In addition to assessing performance for sample discrimination, tests related to online demonstration were conducted at the VTT X-ray laboratory. A successful demonstration of the real-time data fusion was conducted at a speed of 0.24 m/s. Both cameras were configured to operate on a single synchronization signal, and data fusion was implemented by prior image registration. A ring buffer of images was implemented in the data fusion software to account for the physical offset between the cameras. The result is shown in Figure 8.7, where the algorithm output for single channel and stereo vision camera is visible. Two samples appear similar, but due to thickness compensation the other sample is rejected.

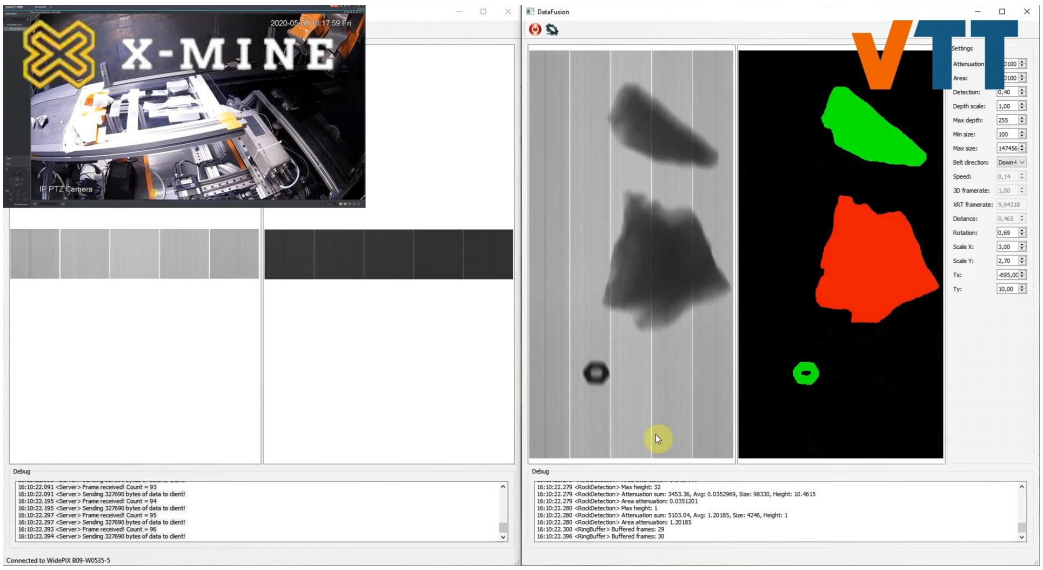


Figure 8.7: Real-time data fusion test at the VTT laboratory. Green and red colors refer to accepted and rejected sample data, which is relayed to pneumatic ejectors in the final application. A slight vertical blurring is present due to mismatch between actuator speed and Widepix integration time.

Chapter 9

Summary and discussion

In summary, relevant X-ray physics and measurement principles were discussed, leading up to the Medipix3 detector for X-ray imaging. Augmentation of X-ray data was achieved by the addition of a stereo vision camera, supplementing complementary information in the form of sample thicknesses. Furthermore, calibration methods for both imaging modalities were discussed. Data fusion was demonstrated and two algorithms for sample discrimination were outlined. In addition, a method for algorithm performance evaluation was introduced. Sorting performance was evaluated for three different mines, and good results were obtained for Lovisagruvan and Hellas Gold mines. However, due to elemental similarity and material distribution within samples, results for Asarel Medet were inadequate. Other imaging modalities need to be considered for improving performance in this case.

Taking into account that stereo camera is considerably less affected by photon noise as sample speed is increased, we can conclude that stereo vision approach yields superior performance in terms of sample throughput. Therefore, the stereo camera is not a limiting factor for the sample throughput or conveyor belt speed. In contrast, the Medipix3 data processing limits the conveyor belt speed, depending on the measurement mode used. Use of dual channel doubles the amount of data and thus, doubles the readout time. At conveyor belt speed 1.5 m/s, theoretical maximum for integration time is 9.4 ms, assuming zero readout time and no geometrical magnification. Considering that integration time for the second campaign was only 1.0 ms, almost 90% of the available integration time is spent by the detector processing the data. Single channel approach halves the amount of data and therefore presumably halves the amount of time spent processing, enabling higher integration time. One can also consider the effect of X-ray source selection; increasing the tube current would yield a higher irradiance and increase the count rate, negating the effects of low integration time. However,

this likely comes at the cost of a larger focal spot size, reducing image sharpness and the benefits of high resolution imaging become questionable. On the other hand, one could simply increase the number of X-ray sources and detectors to increase the irradiance. Furthermore, a higher X-ray source power can enable further threshold optimization. Additionally, one has to take into account the conditions of a higher X-ray flux, where subsequent photon interaction events increase in frequency. The detector is limited in processing speed, and a phenomenon called pulse-pileup can occur [29], where multiple photons are interpreted as a single event due to simultaneous coincidental interactions.

Figure 9.1 demonstrates the principle for integration time between single and dual channel mode, as a function of conveyor belt speed. Equivalent signal-to-noise ratio in terms of photon noise is achieved at much higher speeds when using the single channel mode, enabling much higher sample throughput.

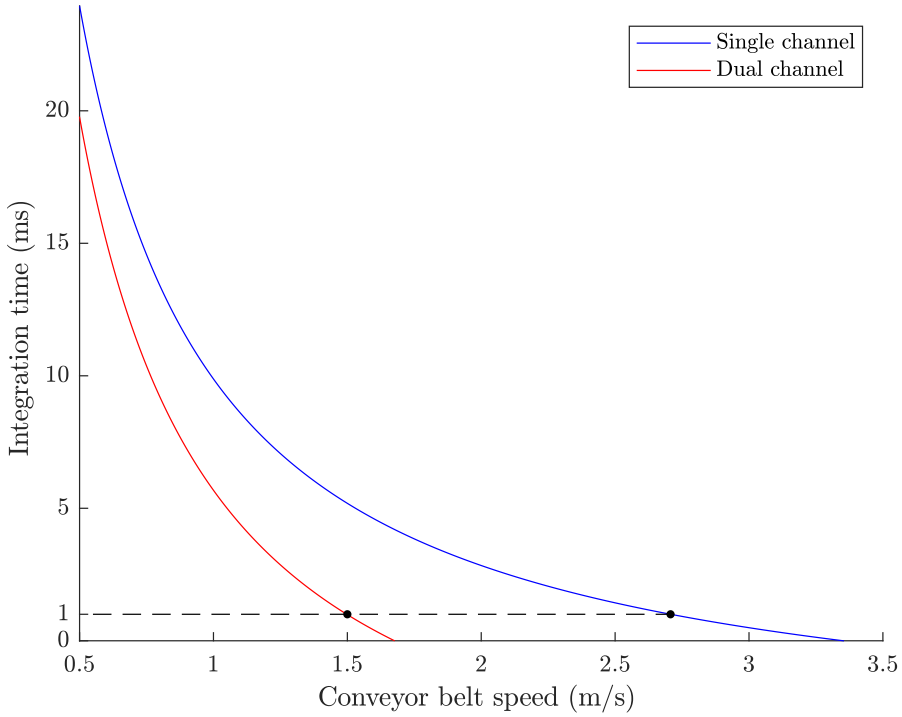


Figure 9.1: Example of integration time as a function of conveyor belt speed for both single and dual channel mode. A lower readout time due to reduced data significantly affects the total sample throughput between the measurement modes.

The performance of the demonstrated system can be improved by recognizing the fact that the samples emit characteristic X-rays specific to the elements within the sample. These characteristic rays can be captured by means of X-ray spectrometers. An array of collimated spectrometers could be employed in such a way as to yield both spatial and spectral information about the samples. As an example, we can consider a sample consisting of mostly low-Z material which encapsulates a valuable high-Z material with a hard X-ray K edge (such as gold). Characteristic K edge X-rays arising from the high-Z material are weakly affected by the low-Z material, yielding volumetric information about the sample composition in the process. Especially in the case for Asarel Medet, use of X-ray fluorescence detection can help distinguish between iron and copper. Additionally, information from other wavelength ranges, such as optical or near-infrared, can be introduced by means of hyperspectral imaging to provide complementary information about the samples. These methods likely only provide information from the sample surface due to shallow penetration depth at the wavelength ranges specified.

It is relevant to note that the thickness estimation described cannot provide any information about the underside of the samples simply due to the measurement geometry. In practice, an upper bound for the thickness is computed. Alternative methods or additional cameras from different viewpoints are needed to increase estimation accuracy. In addition, samples with highly reflective surfaces can produce spurious depth estimations as the cameras are positioned differently. One camera can see a reflection of the illumination and the other likely will not, complicating the stereo correspondence search and reducing the effectiveness of the thickness compensation in performance. One can also consider a case where sample thickness distribution is uniform, and no benefit is derived from thickness measurement, effectively reducing to the baseline performance.

Single channel performance can be improved even further by the use of charge summing mode in Medipix3. Charge summing mode requires the selection of two energy thresholds. Signals that exceed the lower threshold are summed together with neighbouring pixels, and the reconstructed signal is compared to the higher threshold. This reduces the effect of charge sharing, but fluorescence events can still affect imaging performance. The readout time is similar to single channel mode, and no apparent downside to the performance is to be expected.

In conclusion, this thesis demonstrated the operating principles and results of a stereo vision augmented Medipix3 array for real time sample discrimination at high speeds. Using simple reasoning, the benefits of a single channel and thickness measurement should become even more apparent as

the sample thickness variation increases. Further work includes comparisons to existing technologies at pilot scale, where the benefits of high-resolution imaging should manifest.

References

[1] European Commission, *Raw Material Scoreboard*, Publications Office of the European Union, 2016.

[2] X-MINE Project, *Real-time mineral X-ray analysis for efficient and sustainable mining.* `https://cordis.europa.eu/project/id/730270`, 2017. [Online; accessed 15-April-2021].

[3] P. Russo, *Handbook of X-ray Imaging: Physics and Technology*, CRC Press, 2017.

[4] J. Punnoose, J. Xu, A. Sisniega, W. Zbijewski, and J. Siewerdsen, *Technical note: Spektr 3.0 - a computational tool for X-ray spectrum modeling and analysis*, Med. Phys., 43 (2016), pp. 4711–4717.

[5] G. F. Knoll, *Radiation Detection and Measurement*, John Wiley and Sons, New York, 3rd ed., 2000.

[6] A. Compton, *A quantum theory of the scattering of X-rays by light elements*, Physical Review, 21 (1923), pp. 483–502.

[7] R. Ballabriga, *The design and implementation in* 0.13 µm *CMOS of an algorithm permitting spectroscopic imaging with high spatial resolution for hybrid pixel detectors*, PhD thesis, CERN, 2009.

[8] M. Campbell, E. Heijne, G. Meddeler, E. Pernigotti, and W. Snoeys, *A readout chip for a 64 x 64 pixel matrix with 15-bit single photon counting*, IEEE Transactions on Nuclear Science, 45 (1998), pp. 751–753.

[9] X. Llopart, M. Campbell, D. San Segundo, E. Pernigotti, and R. Dinapoli, *Medipix2, a 64k pixel read out chip with 55 Î¼m square elements working in single photon counting mode*, IEEE Nuclear Science Symposium and Medical Imaging Conference, 3 (2001), pp. 1484–1488.

[10] R. Ballabriga, M. Campbell, E. Heijne, X. Llopart, L. Tlustos, and W. Wong, *Medipix3: A 64k pixel detector readout chip working in single photon counting mode with improved spectrometric performance*, Nucl. Instrum. Methods Phys. Res. A, 633 (2011), pp. S15 – S18.

[11] S. del Sordo, L. Abbene, E. Caroli, A. Mancini, A. Zappettini, and P. Ubertini, *Progress in the development of CdTe and CdZnTe semiconductor radiation detectors for astrophysical and medical applications*, Sensors, 9 (2009), pp. 3491–3526.

[12] NIST, *XCOM: Photon cross sections database*.

[13] D. Pennicard and H. Graafsma, *Simulated performance of high-z detectors with medipix3 readout*, Journal of Instrumentation, 6 (2011).

[14] E. Frojdh, *Hybrid pixel detectors: Characterization and optimization*, PhD thesis, Mid Sweden University, Department of Electronics Design, 2015.

[15] D. Wu, L. Zhang, X. Zhu, X. Xu, and S. Wang, *A weighted polynomial based material decomposition method for spectral X-ray CT imaging*, Phys. Med. Biol., 61 (2016), pp. 3749–3783.

[16] T. Schmidt, K. Zimmerman, and E. Sidky, *The effects of extending the spectral information acquired by a photon-counting detector for spectral CT*, Phys. Med. Biol., 60 (2015), pp. 1583–1600.

[17] J. Jakubek, *Data processing and image reconstruction methods for pixel detectors*, Nucl. Instrum. Methods Phys. Res. A, 576 (2007), pp. 223 – 234.

[18] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, 2002.

[19] J. Heikkila and O. Silven, *Four-step camera calibration procedure with implicit image correction*, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., (1997), pp. 1106–1112.

[20] Z. Zhang, *A flexible new technique for camera calibration*, IEEE Trans. Pattern Anal. Mach. Intell., 22 (2000), pp. 1330–1334.

[21] J. Heikkila, *Geometric camera calibration using circular control points*, IEEE Trans. Pattern Anal. Mach. Intell., 22 (2000), pp. 1066–1077.

[22] E. Trucco and A. Verri, *Introductory techniques for 3D computer vision*, Prentice Hall, USA, 1998.

[23] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*, " O'Reilly Media, Inc.", 2008.

[24] A. Fusiello, E. Trucco, and A. Verri, *Compact algorithm for rectification of stereo pairs*, Machine Vision and Applications, 12 (2000), pp. 16–22.

[25] OpenCV, *Open Source Computer Vision Library*, 2020.

[26] H. Hirschmuller, *Stereo processing by semiglobal matching and mutual information*, IEEE Trans. Pattern Anal. Mach. Intell., 30 (2008), pp. 328–341.

[27] M. Fischler and R. Bolles, *Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography*, Communications of the ACM, 24 (1981), pp. 381–395.

[28] A. Goshtasby, *2-D and 3-D Image Registration: For Medical, Remote Sensing, and Industrial Applications*, John Wiley and Sons, 2005.

[29] S. Jorgensen, A. Vercnocke, D. Rundle, P. Butler, C. McCollough, and E. Ritman, *Evaluation of a photon counting Medipix3RX CZT spectral X-ray detector*, Proc. SPIE Int. Soc. Opt. Eng., 9969 (2016).