Lund, January 2016

Master's Thesis

Heterogeneous Multi-Sensor Camera

by

Ida Arvidsson

Johannes Henriksson

Supervisors:
Cord Arnold, LTH
Ola Synnergren, Axis Communications

Department of Physics, Division of Atomic Physics
Faculty of Engineering, Lund University
SE-221 00 Lund, Sweden

Abstract

Today's security cameras typically capture either color images based on visible light or grayscale images based on both visible and near-infrared (NIR) light. This means that the performance is limited in situations where it is unclear which kind of image to capture. At dusk, for example, the visible light may not be sufficient to capture good-quality color images. On the other hand, capturing grayscale images at dusk means that useful color information is disregarded. Thus, there is a need for a hybrid-mode camera which can utilize the best of both the color and the grayscale images. In order to build a prototype camera which can achieve just this, a dichroic beam-splitter was placed in between the lens and the image sensor of a conventional camera. This way, the incident light is split up into its visible and NIR components which then can be detected separately by two image sensors. Different approaches on how to merge color and grayscale images into one superior image were investigated, where the basic idea was to extract the overall intensity from the color image and partly replace it with the intensity of the grayscale image. The prototype camera developed proved to outperform conventional security cameras in certain situations, such as in low light conditions as well as when the illumination is highly varying. While more development is needed, the technique looks promising overall and should offer new imaging capabilities to the security camera market.

Acknowledgements

This thesis was carried out in cooperation with Axis Communications, a network camera company based in Lund. Axis Communications has provided us with supervisors, hardware and workspace which all was absolutely crucial for the success of this project. It has been thrilling to work together with an R&D intensive, fast-growing company and to know that our thesis work possibly could be used in real-world products.

The two persons we have worked closest with have been our supervisors from Axis Communications Ola Synnergren and Sebastian Fors. They came up with the thesis idea in the first place and have helped us carrying out this project by giving support and new input throughout the whole project.

We were part of the Fixed Box Firmware group at Axis Communications and everyone here has been very helpful and interested in our project. We would also like to thank Jonas Hjelmström, optics expert at Axis Communications, for helping us with the optical simulation.

Finally we would like to thank the Division of Atomic Physics at Lund University and in particular Cord Arnold, who has been our supervisor from the university. Cord has always given us useful feedback and it has been very valuable to have him as an extra resource to discuss various problems with.

Popular Scientific Summary: Heterogeneous Multi-Sensor Camera

Imagine a camera which can capture color images using visible light at noon, grayscale images based on both visible and near-infrared light at midnight and hybrid images in between those extremes. Focused on the needs of security cameras, we have built a camera which can do just this.

Today's surveillance cameras typically have one day-mode and one night-mode. In day-mode, the camera captures images from the visible light just like a conventional camera does (see image 1). In night-mode, on the other hand, both visible and near-infrared light is captured which leads to a brighter but colorless image (see image 2).

Image 1. Day-mode image. Too dark to the left.

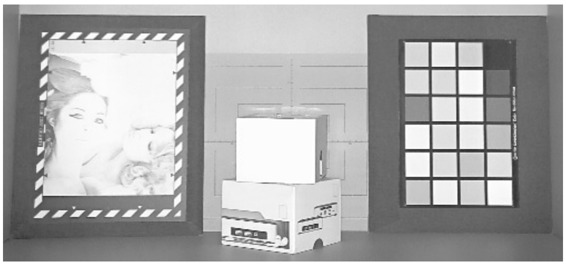

Image 2. Night-mode image. Great contrast.

In situations where it is unclear which mode today's cameras should be in, we believe our camera will be superior. At dusk, for example, the day-mode may not capture enough light to get good-quality images while the night-mode will not capture any color information at all. Another interesting situation is when a street-lamp only illuminates half of the scene which means that an ideal camera would have half of the image in day-mode and the other half in night-mode. None of these situations are covered well by today's camera design which led us to instead look into a fundamentally different design.

From the outside, our camera looks like any other camera with a lens letting the light in and a connection which can transfer the images to computers, TV screens, hard drives or whatever that may be feasible. But when taken apart, our camera has two new elements that a conventional camera lacks. First, it has a second image sensor and second, it has a special thin bit of glass called beam-splitter. The beam-splitter splits the visible light and the infrared light onto the two separate image sensors. This way, the camera produces two sets of images at the same time and thus obtains both day- and night-mode images simultaneously.

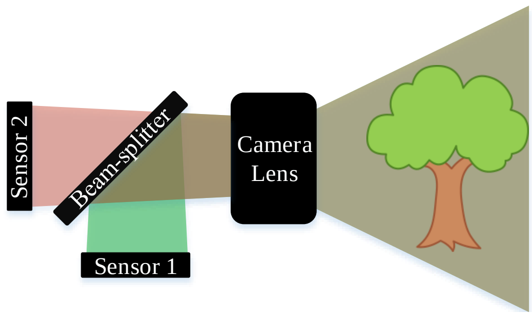

Image 3. The camera setup used.

While this solution of taking two different images from the same camera already is a very interesting idea, we decided to push it all even further and merge the two images into one superior image (see image 4). The merge algorithm first senses how good the two input images are and then essentially extracts the contrast from the night-mode image, the colors from the day-mode image and combines them into a new, merged image.

Image 4. Merged image; the best of image 1 and 2.

Table of Contents

Abstract . . . I
Acknowledgements . . . II
Popular Scientific Summary . . . III
1 Overview . . . 1
2 Background . . . 2
2.1 Camera Design . . . 2
2.1.1 Camera Lens . . . 2
2.1.2 Image Sensor . . . 3
2.1.3 Beam-splitter . . . 5
2.2 Image Processing . . . 10
2.2.1 Image Signal Processing . . . 10
2.2.2 Spatial Image Transformations . . . 10
2.2.3 Color Spaces . . . 11
2.3 Related Work on Merge Algorithms . . . 14
2.4 Imaging Conditions . . . 15
2.4.1 Security Cameras Specific . . . 15
2.4.2 Rayleigh Scattering . . . 15
3 Method . . . 16
3.1 Camera Design . . . 16
3.1.1 Camera Lens and Image Sensors . . . 16
3.1.2 Beam-splitter . . . 16
3.1.3 Mounting and Alignment . . . 18
3.2 Image Merge Methods . . . 19
3.2.1 Test Scene . . . 19
3.2.2 Simple HSV Merge Method . . . 20
3.2.3 Improved HSV Merge Method . . . 21
3.2.4 YUV Merge Method . . . 22
3.2.5 Choosing Parameters . . . 23
3.3 Additional Image Processing . . . 28
3.3.1 Image Signal Processing . . . 28
3.3.2 Digital Alignment . . . 28
3.3.3 Performance Issues . . . 28
4 Results . . . 30
4.1 Different Beam-splitters . . . 30
4.1.1 Sharpness . . . 30
4.1.2 Light Sensitivity . . . 32
4.2 Additional Images . . . 33
4.2.1 Controlled Lab Scenes . . . 34
4.2.2 Outdoor Scenes . . . 37
5 Discussion and Conclusion . . . 38
5.1 Camera Lens and Image Sensor . . . 38
5.2 Beam-splitter . . . 38
5.3 Merge Algorithm . . . 39
5.4 Resulting Images . . . 39
5.5 Competitive Solutions . . . 40
5.6 Considerations Going Forward . . . 41
Bibliography . . . 42

1 Overview

The surveillance cameras used today typically have one day-mode and one night-mode – they capture either only visible light or both visible and near-infrared (NIR) light depending on the current illumination of the scene. The goal with this thesis was to design a camera which captures two images simultaneously, one color image using the visible light and one grayscale image using both the visible and the NIR light, and then merges the images to a superior hybrid image. The wavelength of the NIR light is 750-1200 nm.

This thesis can roughly be divided into two major parts; camera hardware design and image merge algorithms. For the camera design, the basic idea is to insert a 45° tilted beam-splitter right after the lens so that the visible and NIR light is split onto two separate image sensors – see figure 1.1 which illustrates the principle. In the background section of this report different camera elements such as lens, image sensor and beam-splitter are briefly described. In the method section the parts used for building the prototype camera are stated and explained.

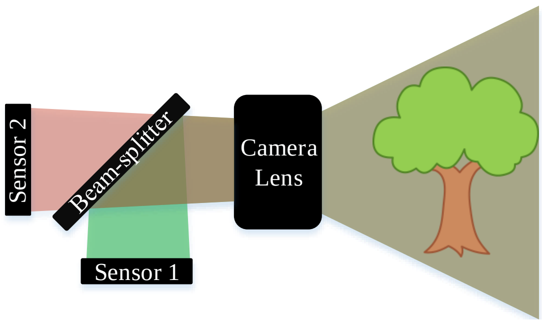

Figure 1.1: The idea of the heterogeneous multi-sensor camera featuring camera lens, beam-splitter and two image sensors. The beam-splitter reflects the visible light (green) onto sensor 1 and transmits the NIR light (red) onto sensor 2.

The second major part of this thesis is about the image merge algorithm which essentially extracts the contrast from the color image and partly replaces it with the grayscale image – see figure 1.2. In the background section general information about, for example, color spaces and spatial transformations is given. In the method section, the information given in the background section is used to describe the algorithms that were developed. In addition, some other software features and considerations are mentioned, such as digital alignment and performance issues.

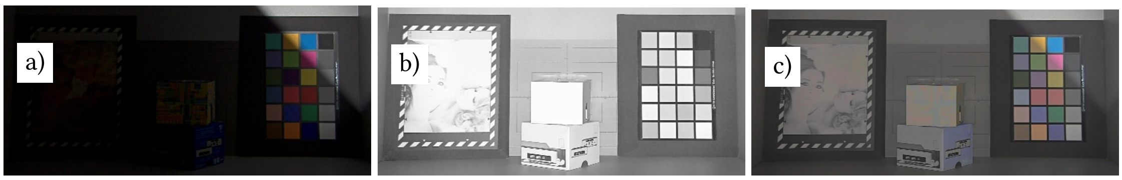

Figure 1.2: Concept of the image merge algorithm: color image (a) and grayscale image (b) are digitally combined into a superior merged image (c).

The results section presents images captured using different camera setups and test scenes. Also, results from an optical simulation of the beam-splitter are included. Finally, in the discussion section, the results are evaluated and the potential of this solution is estimated for security applications.

2 Background

In order to make the rest of the thesis easier to follow, some basic concepts are introduced here. First, key elements used in the developed camera will be described. Then, a brief introduction to image processing is given. Finally, some security camera basics are described.

2.1 Camera Design

The key parts of a conventional camera are the camera lens and the image sensor. These components are described below. Additionally, the concept of a beam-splitter is also introduced, since it will be a key part in the multi-sensor setup.

2.1.1 Camera Lens

The very first element that the light goes through in a camera is the camera lens, which consists of several internal lens elements. Its purpose is to collect the light from the imaging target and focus it onto the image sensor, see figure 2.1. As the distance to the target changes, either the lens will need to be reconfigured or the distance between the lens and the sensor will need to be changed. Otherwise, the camera lens will not be able to focus the incident light properly meaning that the image will appear blurry. How well a camera lens focuses light depends on the wavelength of the incident light. Security cameras which capture both visible and NIR light simultaneously during nighttime have to have a camera lens which focus the visible and the NIR light at the same position. Lenses designed to focus light with different wavelengths to the same focus plane are generally called achromatic lenses and the camera lenses used in this project, designed for visible and NIR light, were labeled IR-corrected.

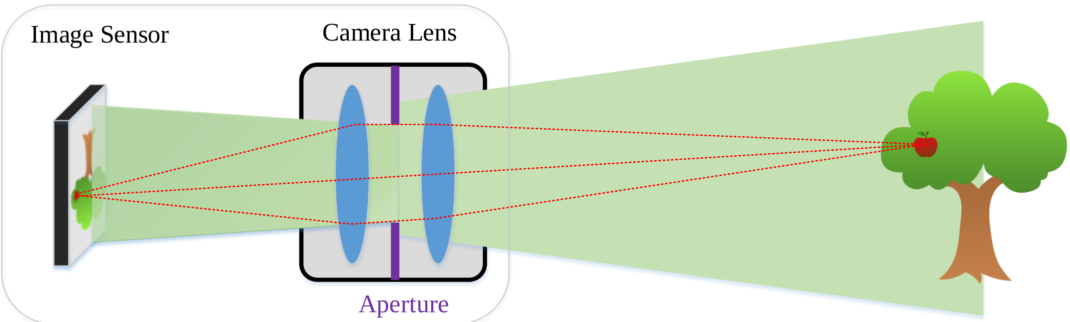

Figure 2.1: The principle of a conventional camera. Note that an actual camera lens consists of more than two lens elements but two were sufficient to show the function of the aperture. As the aperture contracts, i.e. the distance between the two purple lines is decreased, less light is allowed to enter the camera.

One feature of the camera lens is that it can maintain different levels of zoom. Just like focusing, this can be achieved by mechanically changing distances between internal lenses of the camera lens. Once again, it is important to use an IR-corrected lens so that both the visible and the NIR light is focused onto the same plane, even for different levels of zoom.

Another camera lens parameter that was important in this project was the distance from the back of the camera lens to the image sensor. This distance was crucial since the beam-splitter had to fit in between the camera lens and the image sensor. Additionally, it was important that changes in lens configuration, such as changing the zoom or focus, did not mean that the outer size of the camera lens changed significantly.

Yet another important feature of a camera lens is its aperture, which refers to an opening with changeable size through which all of the outgoing light will have to pass, see figure 2.1. For large apertures, a lot of light passes the opening and is focused onto the sensor – often resulting in a brighter image. Generally, a small aperture means that the depth of field is longer, i.e. that objects at a wider range of distances will be focused and not blurred. As will be discussed later, depending on the thickness of the beam-splitter, it may be beneficial to use small apertures in order to get sharp images.

All lenses give rise to spherical aberration, which means that light hitting the lens at different positions will be focused at different depths. For a camera lens this leads to that an image never can be focused perfectly. Without any spherical aberration the resolution would be limited by the airy disk, but the effect from the spherical aberration usually is much larger. By decreasing the maximal allowed angle of incidence, typically by decreasing the aperture, this effect can be decreased and sharper images are obtained. Spherical aberration is also obtained if a plate of glass is placed after the camera lens. Since many camera designs contain a piece of glass between the camera lens and the image sensor, the camera lenses usually are designed to give the least spherical aberration for a specific amount of glass. In this project this is desirable, since a beam-splitter will be placed between the camera lens and the image sensors [1].

For a thin convex lens the focal length is the distance at which collimated light is focused, see figure 2.2. For a camera lens, which is built out of multiple single lenses, the focal length definition is more complicated but the idea is still the same. The ratio between the focal length and the aperture diameter is often specified instead of the aperture size itself. This unitless ratio is called the camera lens's f-number and a low f-number means that much light is let into the lens, better resolution and smaller depth of focus.

Figure 2.2: A convex lens of diameter d focuses parallel light to the focal point. The distance between the lens and the focal point is called the focal length f.

2.1.2 Image Sensor

The image sensor can be regarded as the heart of a digital camera since it detects the light and outputs an image as a digital signal. There are two common types of image sensors; the charged coupled device (CCD) and the complementary metal oxide semiconductor (CMOS). Here, only CMOS image sensors will be described since they were used in this project. However, both types of image sensors share many features.

In order to capture the spatial variation of an image, the image sensor is built up by typically millions of individual pixels, each with an area of a few μm^2. In conventional image sensors each pixel is detecting either red, green or blue light. This is achieved by putting a color filter array in front of the pixels. The most common design pattern is called the Bayer filter and it is shown in figure 2.3 [2].

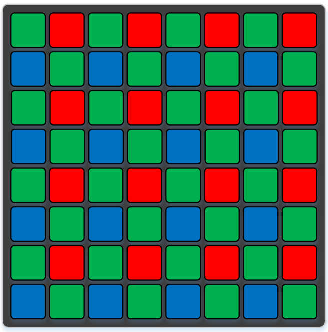

Figure 2.3: The design pattern of the Bayer filter. Note that green light is allowed to be transmitted two times more often than red and blue light, respectively. Green is chosen because the human eye is most sensitive to green light [2].

The light sensitivity for different sensors varies, but a typical relative sensitivity is given in figure 2.4. Since all pixels are sensitive to the NIR light, the grayscale image can be captured with the same kind of sensor. Note that the human eye is not sensitive to NIR light and thus the developed camera ultimately obtains a broader spectral detection range than the human eye.

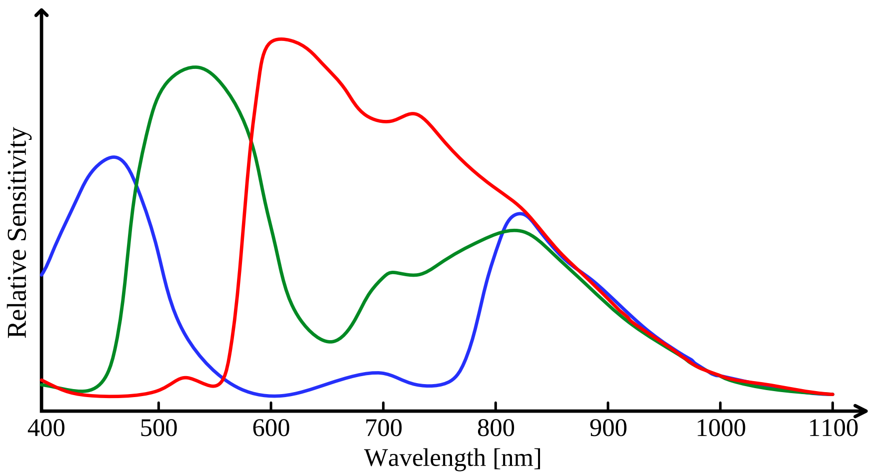

Figure 2.4: The sensitivity for blue, green and red pixels. Note that all kinds of pixels are sensitive in the NIR part of the spectra around 800 nm.

The fundamental physical process when light is converted to an electrical signal in a single pixel can be understood through semiconductor physics, see for example chapter 18 in Fundamentals of Photonics [1]. While the exact design may vary, the basic idea is the same as in a photodiode. When a photon with sufficient energy hits the depletion region it is absorbed and creates an electron-hole pair. The charges, i.e. the electron and the hole, are then separated and a photocurrent is created [3].

The electron charges created by the incident photons are stored in a potential well for a certain period of time, called the exposure time. Then the well is read out as a voltage and reset. Also, a gain may be applied to increase the signal. All of this is done for every pixel – an important feature of the CMOS image sensor. There is also room to do some image processing on the sensor chip, more about this later in the image signal processing part [3, 4].

The gain and the exposure time are two of the key parameters of an image sensor. If an image is too dark, there are two ways to brighten it; increase the gain or increase the exposure time. If the gain is increased, the noise will also be amplified leading to a more noisy image, see figure 2.5. If the exposure time is increased, the highest possible frame rate, if in video mode, will be decreased. Also, the camera will be more sensitive to motion blur if the exposure time is high. Capturing good images in different situations is thus often a balancing act between finding good values for the gain and exposure time.

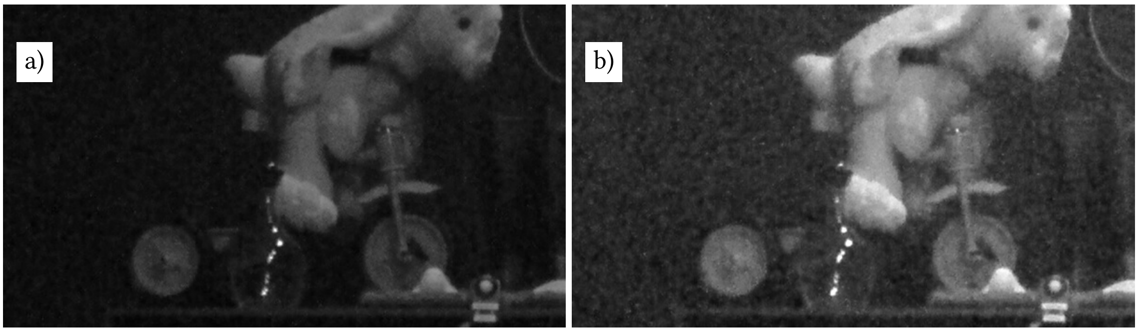

Figure 2.5: The effect of the gain on the sensor; a lower gain gives a darker image (a), while a higher gain gives a brighter but more noisy image (b).

2.1.3 Beam-splitter

The main difference between the multi-sensor camera setup and a conventional camera is the beam-splitter. This element makes it possible to divide the incoming light onto the two image sensors, see figure 2.6. Two different kinds of beam-splitters have been tested; one dichroic glass-based beam-splitter and one pellicle beam-splitter and they are both described below. The beam-splitter used will introduce blur to the transmitted image. The effect is investigated through calculations and simulations.

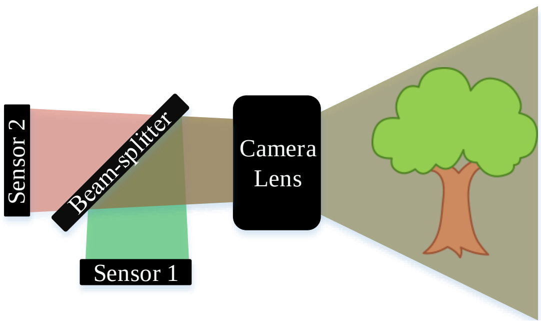

Figure 2.6: The dichroic beam-splitter reflects the visible light (green) onto sensor 1 and transmits the NIR light (red) onto sensor 2.

An ideal beam-splitter would reflect light only at its first surface. However, some light will be reflected at the second surface too and give rise to a so-called ghost image. This image is undesirable since it is spatially shifted compared to the image from the first reflection. Also, the ghost image will not be focused in the same plane as the main reflection since this light will have traveled a longer distance. The same phenomenon can occur for the transmitted image if it is reflected twice in the beam-splitter.

If the beam-splitter would have been inserted at 0° it would give rise to spherical aberration, because light hitting it with different angles will be focused at different focal planes. Thus it will not be possible to focus the light properly afterwards. The phenomenon is dependent of the thickness of the beam-splitter, and a thinner beam-splitter gives rise to less spherical aberration. If the beam-splitter instead is tilted 45°, as will be the case in this setup, the phenomenon will be similar and the thinner beam-splitter will still give less aberration. Note that camera lenses often are designed to give minimal spherical aberration for a given amount of glass inserted after the camera lens [1].

A dicroic beam-splitter is coated so that light hitting it will be reflected or transmitted depending on the light's wavelength. For the multi-sensor camera, this type of beam-splitter was ideal, since optimally the visible light should be separated from the NIR light. The coating consists of multiple thin layers with different refractive indices and thicknesses. Each layer gives constructive interference for the wavelengths for which the reflection of the second surface is in phase with the reflection of the first surface, see figure 2.7 where n_i are the refractive indices of the different layers.

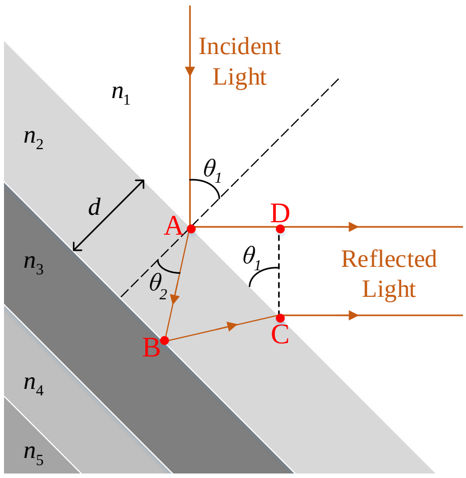

Figure 2.7: The principle of a coated dichroic beam-splitter. The thickness and refractive index of a layer (gray) can be adjusted so that it reflects a specific wavelength.

Constructive interference is obtained if the difference between the optical distances $n_2(\overline{AB}+\overline{BC})$ and $n_1\overline{AD}$ is equal to a multiple of the wavelength. By using the geometry in figure 2.7 and Snell's law, see equation (1), the wavelengths for which this condition holds can be determined, see equation (2) [1].

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{1}$$

$$k\lambda = 2dn_2 \cos\theta_2, \quad k \in \mathbb{N} \tag{2}$$

By having multiple layers it is possible to construct a broadband dichroic beam-splitter and split off light with wavelengths within a desired range. The principle for the next layers is the same as for the first, but the calculations become more complicated as the number of possible reflections increases. Furthermore, a π radians phase shift is introduced for the reflections from the surfaces where the refractive index of the next layer is higher than the refractive index in the previous layer. Often only two different materials are used, thus the refractive indeces alternate between n_1 and n_2, and only the thicknesses of the layers are varied to reflect different wavelengths. A multilayer coating is typically put on a glass plate, which often has an anti-reflective coating on its other side. This is to reduce undesired reflections which otherwise would give rise to ghost images. The thickness of this type of glass beam-splitter is about 0.2 - 5 mm [5].

The pellicle beam-splitter is a very thin film, about a few micrometers thick, and is therefore very fragile. The film can be coated to adjust how much of the light that is reflected. An advantage of the pellicle beam-splitter compared to the thicker ones is that ghost images are eliminated. This is because the reflection on the second surface is superimposed with the one from the first reflection [6]. Another advantage of the pellicle beam-splitter is its high flatness. The flatness quantifies how much the surface differs from a perfectly plane surface, and a low flatness ultimately leads to more blurry images [7].

A glass plate with an angle to the incident light will give rise to a shift Δx_1 orthogonal to the propagation direction, see figure 2.8. This shift will determine how well the image can be focused and therefore it is important to examine this effect. The underlying reason is that light with different angles of incidence will travel different path lengths in the beam-splitter and therefore be shifted differently. For a beam-splitter with thickness d and refractive index n_{bs} the shift can be calculated by using the geometry in figure 2.8 and Snell's law given in equation (1). The shift Δx_1 is given in equation (3).

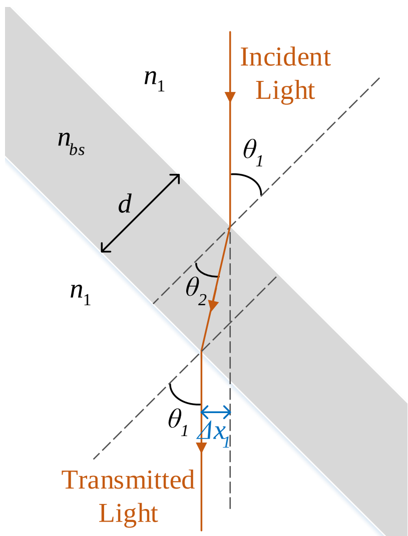

Figure 2.8: This geometry can be used to calculate the orthogonal shift of the light, Δx_1, as a function of the angle of incidence θ_1.

$$\Delta x_1(\theta_1) = d\frac{\sin(\theta_1 - \theta_2)}{\cos\theta_2} = d\sin\theta_1(1 - \frac{\cos\theta_1}{\sqrt{n_{bs}^2 - \sin^2\theta_1}}) \tag{3}$$

The light will on average hit the beam-splitter at $45°$. The difference in shifts $\Omega(\theta_1)$, see equation (4), is given in figure 2.9, for refractive index $n_{bs} = 1.5$ and thickness $d = 0.2$ mm of the glass plate. For the camera lens used in the multi-sensor setup the angles of incidence on the beam-splitter vary between 30° and 60°, and thus the variation of the shift is about 60 µm. Compared with the size of a pixel, about a few µm^2, this shift is large and will definitely affect the captured images.

$$\Omega(\theta_1) = \Delta x_1(\theta_1) - \Delta x_1(45^\circ) \tag{4}$$

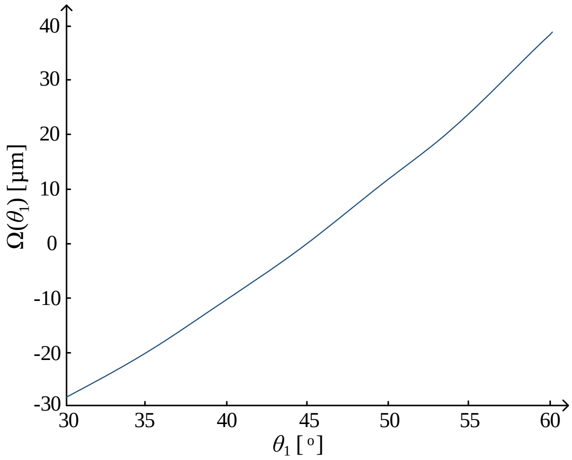

Figure 2.9: The orthogonal shift $\Omega(\theta_1) = \Delta x_1(\theta_1) - \Delta x_1(45^\circ)$ as function of incoming angle θ_1 of the light, when a 0.2 mm thick beam-splitter is used.

Since the shift varies depending on the light's angle of incidence, it will not be possible to focus the images properly after the beam-splitter, see figure 2.10. Since the shift is dependent on the thickness of the beam-splitter, it is important to choose an as thin beam-splitter as possible. Another approach would be to narrow the distribution of the incoming angles on the beam-splitter. This can be done by decreasing the size of the aperture, i.e. by increasing the f-number.

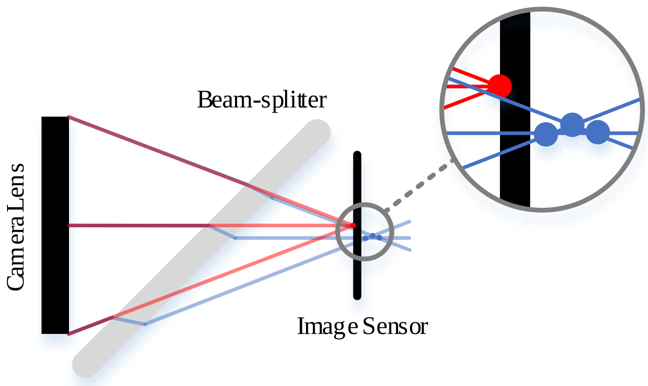

Figure 2.10: Without any beam-splitter (or with a very thin beam-splitter) the camera lens focuses light onto one point on the sensor (red) giving a sharp image. The beam-splitter introduces an angle of incidence dependent shift, making it impossible to focus all rays to one point (blue) leading to a blurry image.

We have not found any optical component that can fully compensate for the shift. One could imagine adding a second glass plate with the same refractive index n_{bs} as the beam-splitter and with an angle of 90° to the beam-splitter. Then the total shift would cancel for light with an incidence angle of $\theta_1 = 45^{\circ}$ but the shift would increase for other angles and thus this would not solve the problem.

2.2 Image Processing

In order to be able to merge the color and the grayscale images together, they first need to be very well aligned. In order to achieve this, some different spatial image transformations were examined, which are described below. Then a couple of color spaces are described, which were used to combine the color and the grayscale images. Some of the previously developed merge algorithms were also investigated, which are briefly described. Those were used for inspiration when the merge algorithm optimized for security cameras was implemented. First, however, some basic image signal processing is described.

2.2.1 Image Signal Processing

There is an increasing number of operations that are applied and can be applied to images before the user even sees the image. Some are applied on sensor level, others in the image processing pipeline. Of course the processing also can be done offline, once the images already have been read out from the camera. Here, some of the key parts of image signal processing (ISP) will be mentioned, especially those that have been relevant to this project.

Demosaic refers to the process where an image's pixels are assigned three values each for the level of red, green and blue light. As explained previously, each pixel only measures either the level of red, green or blue. For a given pixel, the demosaic algorithm essentially looks at its neighboring pixels with different color and estimates the level of red, green and blue light by interpolation.

A white balancing algorithm changes the gain of individual color channels. It is needed since different environments have different illuminations, and to get correct colors the gain of the different colors need to be adjusted. For example, the illumination indoors typically has another distribution than then illumination outdoors. Actually, due to for example the sensor characteristics, the white balancing algorithm is absolutely crucial in order to get natural looking images in most situations.

Another operation that often is applied is the so-called gamma correction, which converts the linear response of the sensor to a non-linear representation more similar to the way our eyes perceive light. Also, a defect pixel correction is used to limit individual pixel errors originating from the sensor manufacturing process. Another image processing algorithm that should be mentioned is the tone-mapping, which in this project often had the effect of making the output image brighter. It can be applied both globally, so that the whole image gets brighter, or locally so that only dark regions receives an increase in brightness. Finally, there are also different algorithms to reduce noise.

2.2.2 Spatial Image Transformations

Before the color and grayscale images can be merged, they need to be spatially aligned to each other. It is difficult to align the sensors perfectly and also the images will always have slightly different zoom because of the wavelength-dependent camera lens. Therefore a spatial transformation was applied on one of the images, to make it aligned to the other image. The perspective transformation which is very general was used. It is described below. First two simpler transformations, the similarity and the affine transformation, are described. Illustrations of the transformations are given in figure 2.11.

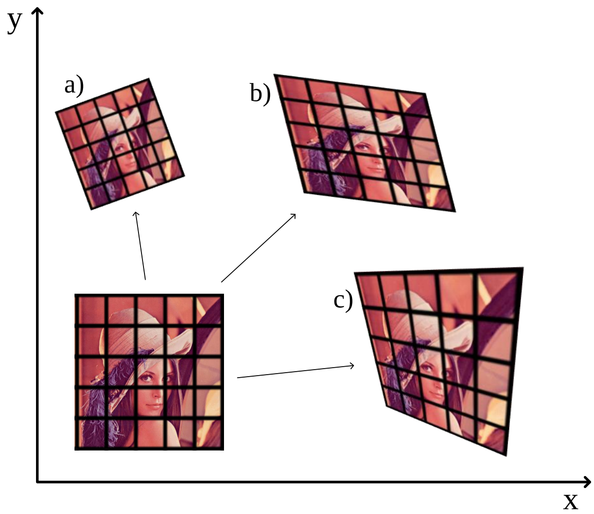

Figure 2.11: Illustration of the similarity transformation (a), the affine transformation (b) and the perspective transformation (c).

Let $\mathbf{x}$ describe the coordinates for each pixel in the image, i.e. $\mathbf{x} = [x\ y]^T$ where x and y are the x- and y-coordinate. To translate the image, a constant should be added to the x- and/or the y-coordinate, i.e. $\mathbf{x'} = \mathbf{x} + \mathbf{t}$ where $\mathbf{t}$ is the translation vector and $\mathbf{x'}$ is the new pixel coordinates. The rotation of an image with an angle θ is obtained if each pixel's coordinates are transformed by $\mathbf{x'} = \mathbf{R}\mathbf{x}$, where the rotation matrix $\mathbf{R}$ is given by equation (5). The total transformation if the image also should be scaled, in addition to translated and rotated, is given by $\mathbf{x'} = s\mathbf{R}\mathbf{x} + \mathbf{t}$, where s is the scaling factor. This transform is known as the similarity transformation, and preserves angles between lines in an image.

$$\mathbf{R}(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{5}$$

The affine transformation allows translation, rotation, scaling and shearing. It can be represented by an arbitrary 2x3 matrix $\mathbf{A}$, and the transform is given by $\mathbf{x'} = \mathbf{A}\bar{\mathbf{x}}$ where $\bar{\mathbf{x}} = [x\ y\ 1]^T$. Using this transform, angles between lines are not longer preserved but parallel lines will remain parallel after the transformation.

In order to compensate for as many differences between the two images as possible, the perspective transformation was ultimately used. This transformation can be represented by an arbitrary 3x3 matrix $\tilde{\mathbf{H}}$. Since this matrix only is defined up to a scale, i.e. two matrices which only differs by a scale are equivalent, the resulting coordinates $\tilde{\mathbf{x}}' = \tilde{\mathbf{H}}\bar{\mathbf{x}}$ have to be normalized. Using this transformation straight lines remain straight, but angles between them can change and parallel lines may not remain parallel under the transformation. For more information about spatial image transformations, see for example chapter 2 in Computer Vision: Algorithms and Applications [8].

2.2.3 Color Spaces

A color image can be represented in different ways by using different color bases, where some bases are more convenient than others for image processing. Three of the most common color spaces, the RGB, HSV and the YUV spaces, are described below.

The most common color space is probably RGB space, which specifies red, green and blue intensities for each pixel. An equal amount of each color corresponds to different gray levels, see figure 2.12. Since most image sensors are designed with a Bayer filter to let in red, green or blue light to different pixels, RGB is a natural choice of color space.

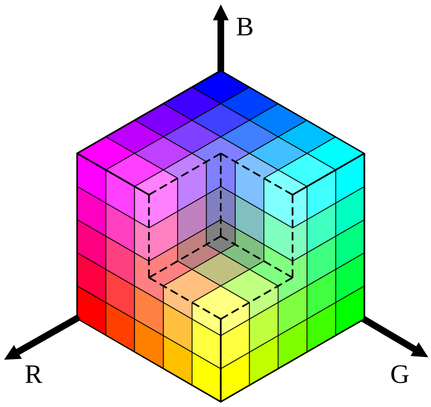

Figure 2.12: Illustration of RGB color space as a cube. The two corners that are not displayed are black (origin) and white (R,G,B all max) [9].

The HSV color space is developed to be intuitive, by approximating how humans perceive and interpret colors. Its three coordinates represent the hue, saturation and value in the image instead of the red, green and blue intensity, see figure 2.13. V is simply the maximum of the R, G and B values and for S = 0 it represents different gray levels, independent of the value of H. If S is decreased more white is added while when V is decreased more black is added [10]. The conversion between RGB and HSV color space is given in equation (6) [11].

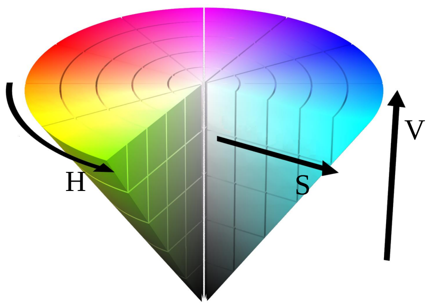

Figure 2.13: Illustration of HSV color space as a cone [12].

$$H = 60^o \cdot \begin{cases} \text{undefined} & \text{if } \max(R, G, B) = 0 \\ \frac{G-B}{\max(R,G,B)-\min(R,G,B)} & \text{if } R = \max(R, G, B) \\ 2 + \frac{B-R}{\max(R,G,B)-\min(R,G,B)} & \text{if } G = \max(R, G, B) \\ 4 + \frac{R-G}{\max(R,G,B)-\min(R,G,B)} & \text{if } B = \max(R, G, B) \end{cases}$$

$$S = \begin{cases} \frac{\max(R,G,B)-\min(R,G,B)}{\max(R,G,B)} & \text{if } \max(R, G, B) \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$$V = \max(R, G, B)$$

By the definition of HSV space, the hue is undefined if V = 0. Also for small values of V, the hue value is very sensitive since all colors appear very similar when they are dark. The same is true for the saturation. This means that the H and S values will be noisy, which is not a problem as long as V is unchanged and low. If, however, V is manually increased, the variation of H and S no longer leads to small but instead large variations in perceived color. This means that the image appears noisy. This will be a problem when the color and grayscale images should be merged together, see the method section.

When processing an image, it is often convenient to have one channel which represents the intensity, i.e. it represents an image similar to the grayscale image based on the NIR light. This can be interpreted as the inverse of what the Bayer filter does, i.e. the intensity should be represented (approximately) by the values obtained if no filter was used. A color space which quite well fulfills this is the YUV color space. Here Y represents the luma, or the brightness, in the image, while U and V give the color representation, see figure 2.14 [13]. Note that this V is not the same as the V channel in the HSV representation. The brightness of the image can be modified by changing the Y-channel, without affecting the observed colors. For example information can be taken from the intensity of the grayscale image, that was captured based on both the visible and the NIR light, and weighted together with the Y-channel.

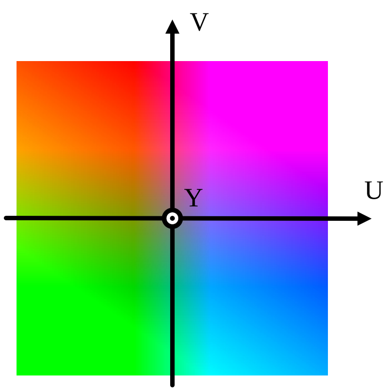

Figure 2.14: Illustration of YUV color space as a cross section of a cube. The Y-axis perpendicular to the plane regulates the luma value [14].

The luma Y is a linear combination of R, G, and B with different coefficients for the different colors; $Y = 0.299R + 0.587G + 0.114B$. The U and V channels are given by $U = 0.492(B\text{-}Y)$ and $V = 0.877(R\text{-}Y)$. Since the conversion between the RGB and the YUV spaces is a linear transformation, it can be expressed with a matrix, see equation (7). The transformation back from YUV to RGB is given by the inverse of this matrix [10].

$$\begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{pmatrix} \tag{7}$$

2.3 Related Work on Merge Algorithms

The idea of capturing images based on either the visible or the NIR light and then combine them is not novel and work on this topic has been done before. Often, such work seems to focus on other areas than real-time security camera applications. Therefore, a relatively simple algorithm more suitable for this application was developed instead. It is described in the method section. However, some research was done before the development of this algorithm, to get inspired by related previous work. A few references that could be useful in the future are listed next.

The algorithm outlined by Zhang et al., which can be used to enhance the quality of photographs [15], was partly implemented but turned out to be too slow. The method is rather sophisticated and includes the Haar transformation, gradient magnitude matching and a weighted region mask. The idea of the weighted region mask was however used in the developed algorithm too, but defined and used in a slightly different way. More specific problems where similar methods can be used include denoising by Honda et al. [16], improve finding SIFT points by Brown et al. [17], skin smoothing by Fredembach et al. [18], removing shadows by Salamati et al. [19] and dehazing by Feng et al. [20]. Generally, these methods seemed complex and included time-consuming and complicated algorithms to solve inverse problems and optimization problems among others. They seemed to be designed for post-processing and not real-time applications; for example it is stated by in Feng et al. that merge two images took them 60 seconds.

2.4 Imaging Conditions

2.4.1 Security Cameras Specific

Security cameras are built to serve in other situations than cameras meant for the consumer market, such as those in mobile phones. The largest difference is perhaps that there often is a night-mode in security cameras. In day-mode, there is an IR-cut filter blocking all infrared light so that it does not reach the sensor and messes up the colors – just like in conventional cameras. In night-mode the IR-cut filter is removed so that both visible and NIR light can be detected by the sensor. This is possible because the image sensors used are silicon based and silicon is sensitive to not only visible light but also to NIR light. The reason why it is beneficial to use night-mode in low light conditions, such as at nighttime, is that then both visible and NIR light is captured and thus the total amount of light is higher. Also, security cameras often have IR LEDs to illuminate the scene during nighttime. Note that this illumination will be invisible to those in the scene – they will believe they act in darkness when there actually is plenty of NIR light available for the camera to capture.

The images captured in night-mode essentially look like a black and white representation of the color images captured in day-mode. Thus a similar image would be obtained if the red, green and blue pixels were added to one value. This is true for most objects, but some objects, for example grass, trees and some dark fabrics, reflect a lot of IR light and thus appear very bright in the grayscale images. This will be a problem when the images should be merged together, which will be discussed in more detail later.

Since the security cameras need to perform well during low light conditions, the image sensor should have good sensitivity. Additionally, the camera lens often has a low f-number to get as much light into the camera as possible. It is often not feasible to use too high exposure time since the images will then get blurry if something is moving in the scene. Furthermore the maximal number of frames per second will be limited since it cannot be higher than the inverse of the exposure time.

2.4.2 Rayleigh Scattering

In certain imaging situations with particles causing haze, smoke or smog, it may be beneficial to capture images based on light with wavelengths outside the visible spectrum. The underlying phenomena causing the degradation in sharpness is Rayleigh scattering, which is strongly wavelength dependent. Thus at certain conditions it can be advantageous to capture images based on both the visible and the NIR light even during daytime.

Rayleigh scattering is an elastic process between light and particles which are much smaller then the wavelength of the light. The particles can be either single atoms or molecules. The probability for Rayleigh scattering is proportional to λ^{-4}, where λ is the wavelength of the light. Hence the probability decreases with increasing wavelength and the infrared light is scattered less than the visible [21].

3 Method

In this section, a description of how the project was carried out follows. First, the camera design is covered, including choice of components such as lens and image sensors. Then, a few different image merge methods are described and visualized. Finally, there is a description of some of the other image processing done in the project.

3.1 Camera Design

3.1.1 Camera Lens and Image Sensors

Since security cameras are used in low light applications it is often desirable to use lenses with low f-numbers. An IR-corrected camera lens from Ricom with f/1.4 (in wide-mode) was used for the prototype camera. It had adjustable focus, zoom and aperture size. Another important feature of the chosen camera lens was that the distance between it and the image sensor was relatively large, about 17 mm depending on the level of the zoom and focus. This was important since there had to be space for the beam-splitter in between the camera lens and the image sensors.

As far as image sensors are regarded, a chip from OmniVision was used to capture both color and grayscale images. The motivation for this choice was not the sensor's performance but rather the size of the printed circuit board (PCB) that the sensor is mounted to. Being able to fit both of the image sensors and the beam-splitter was a big challenge in this project and it automatically put constraints on some parts of the design. It should be mentioned that in a more optimized setup, the choice of image sensors would probably be different. For example, the image sensor set to capture NIR light, could had been specially designed to do just this by, for example, removing its Bayer filter.

3.1.2 Beam-splitter

As mentioned in the background, two different beam-splitters with different advantages were tested. The dichroic beam-splitter splits up the light depending on the wavelength. This means that if the beam-splitter is well-designed no IR-cut filter will be needed in front of the image sensor that is responsible for capturing visible light. On the other hand, a thick dichroic beam-splitter gives rise to orthogonal shifts causing blurry images. This problem can be avoided by instead using a thin pellicle beam-splitter.

The dichroic beam-splitter used is originally designed to act as an IR-cut filter for light hitting it perpendicular, i.e. the visible light is transmitted and the NIR light is reflected. It was taken from a camera developed by Axis Communications. As described in the background, the coating consists of multiple thin layers which give constructive interference for certain wavelengths, the NIR wavelengths in this case. The reflectance of the dichroic beam-splitter at 0° angle of incidence can be seen in figure 3.1.

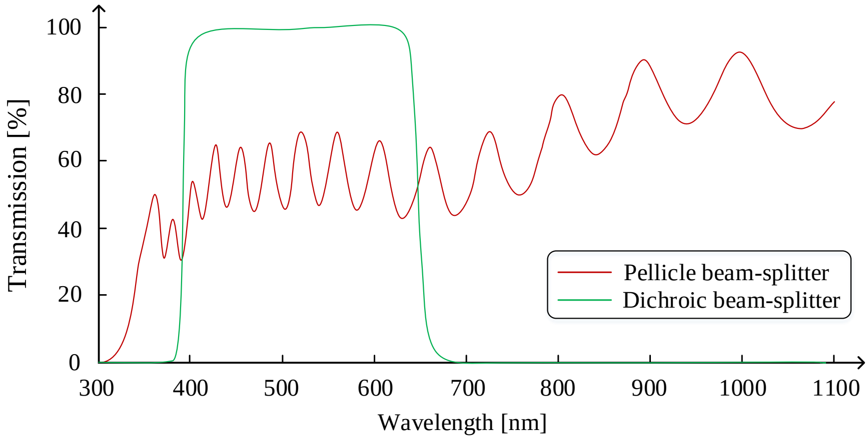

Figure 3.1: Transmission for the dichroic beam-splitter at 0° angle of incidence (green) and for the pellicle beam-splitter at 45° angle of incidence (red). We did not manage to get data for 45° for both beam-splitters which unfortunately makes it impossible to make a completely fair comparison.

If the light's angle of incidence is changed, the cut-off wavelength for the beam-splitter will also change. According to equation (2), the cut-off wavelength will decrease as the angle of incident goes from 0° to 45°. For the dichroic beam-splitter used, the cut-off wavelength changed from about 650 nm to 570 nm as the beam-splitter was tilted to 45°. As a consequence, some of the red visible light is reflected but since still plenty of red light was transmitted, this was never a huge problem. The relatively low percentage of red light compared to green and blue light hitting the sensor could be compensated for digitally to some degree, by using a white balancing algorithm.

The dichroic beam-splitter used is 0.21 mm thick. As explained in the background section, this led to rather large spatial shifts of the light. Therefore, the images taken with this beam-splitter got blurry. In addition to the calculations shown in the background section, simulations were also carried out using Zemax. The results from these simulations are presented in the results section.

The pellicle beam-splitter used was obtained from Thorlabs and it is coated to reflect and transmit equally for wavelengths between 400 and 700 nm, see figure 3.1. For longer wavelengths the transmission is higher than the reflection. Therefore the transmitted image was captured as the grayscale image. The sinusoidal fluctuations of the transmission is caused by interference in the beam-splitter [6].

The pellicle beam-splitter is made of a 2 μm thick nitrocellulose membrane [6]. Since the beam-splitter does not split the light depending on wavelength, it was necessary to insert an IR-cut filter in front of the sensor meant to capture visible light. No IR-pass filter was put to block the visible light hitting the other sensor but doing so would also be a viable design.

The main drawback of using a pellicle beam-splitter is that it does not reflect and transmit the light depending on wavelength. This means that some of the light cannot be utilized and will be thrown away. This is not good, especially not for security cameras which often are used in low-light conditions. Ideally, a dichroic beam-splitter would be used that could both split the light as desired and be thin enough to not introduce any significant blur.

3.1.3 Mounting and Alignment

Once the beam-splitter, camera lens and image sensors had been chosen the next step was to put everything together. To do this, a big box designed to build prototype cameras in was used as the base of the camera. Inside this box, the two image sensors were mounted on one 6-axis alignment stage each. This meant that they could be translated along all three axes as well as rotated it about the same three axes. The camera lens was attached to a threaded hole in the box and the size of the aperture could be changed electronically while focus and zoom could be changed manually. The beam-splitter was mounted in a rigid way without any possibility to easily shift or rotate it. In figure 3.2 the final setup, where the pellicle beam-splitter was used, is displayed.

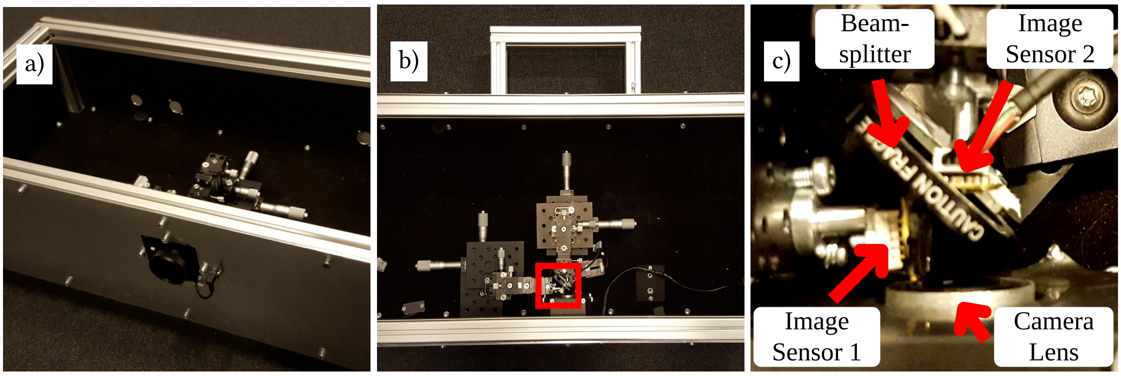

Figure 3.2: The setup with a pellicle beam-splitter built in a rigid box. The front (a), from above (b) and zoomed in from above (c).

A significant amount of time was dedicated to aligning different versions of the system. It was relatively easy to center the image sensors but more tricky to change the tilt of the sensors in order to get an even level of sharpness in the images. The strategy used for the alignment process was to keep the camera lens and beam-splitter in the same place while changing the position and rotation of the image sensors. This meant that the degrees of freedom could be reduced but also that the beam-splitter's position most likely was not ideal neither tilted exactly 45°.

Another challenging aspect was that everything had to be mounted extremely tightly. The resources to use customized beam-splitters or sensor boards were insufficient, and instead the smallest and most suitable parts available had to be used. This implied frequent problems with centering the image sensors and limitations in focusing the image at certain zoom levels.

3.2 Image Merge Methods

The goal with the image processing algorithm is to take one color image and one grayscale image and combine them to one superior image, where all features obtained with both the visible and the NIR light contribute to the final result. Below, steps of different merge algorithms are illustrated, all using the same test scene.

3.2.1 Test Scene

During the development of the merge algorithm another camera from Axis Communications with both day-mode (IR-cut filter on) and night-mode (IR-pass filter on) was used. This made it possible to test the algorithm before the prototype camera was built. Since the images were captured with the same camera it was not necessary to align the images to each other. However, a static scene was needed since the images could not be captured simultaneously.

The scene used can be seen completely illuminated in figure 3.3. The illumination conditions in the scene were changed so that the color and grayscale images contained different information. Both visible light from a flashlight and IR light from Axis Communications IR LEDs were used to change the light conditions. The IR light illuminates the whole scene while the visible light only partly illuminates the scene which makes the scene realistic for security cameras. The resulting illumination of the scene which was used to test the different merge algorithms can be seen in figure 3.4.

Figure 3.3: Full illumination of the test scene.

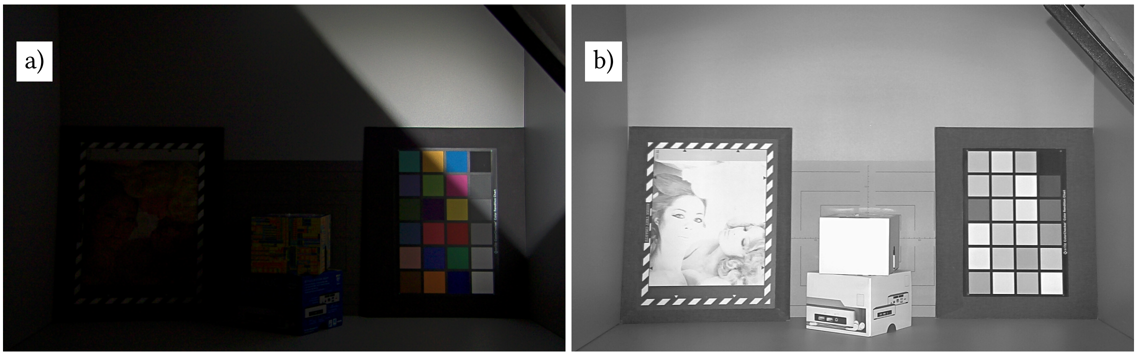

Figure 3.4: Original color image (a) and grayscale image (b) used for the development of the image merge method.

3.2.2 Simple HSV Merge Method

The first approach to merge the images was by transforming the color image from RGB color space to HSV color space and then simply replace the V channel with the intensity of the grayscale image. The H, S and V channels are given in figure 3.5. The V channel is replaced with the image in figure 3.4 b), and the result of this can be seen in figure 3.6.

This method has some drawbacks and the result looks rather noisy, in particular the painting on the left side of the image. The painting is very dark in the original color image meaning hardly any color information can be extracted. The resulting image is very noisy, see figure 3.6. Note that the painting is rather bright in the grayscale image, see figure 3.4 b), which means that the resulting merged image also will be bright here. Therefore, the noisy color information in the painting will be amplified and the result is consequently not very satisfying.

Another problematic case with this particular method is when objects that reflect IR light very well, for example green trees, are imaged. The tree brightness will then be increased a lot resulting in the tree getting an unnatural bright shade of green.

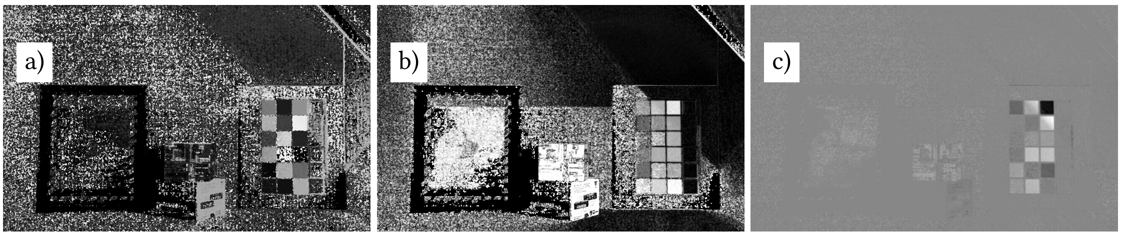

Figure 3.5: Hue (a), saturation (b) and visibility (c) channels in the HSV color space.

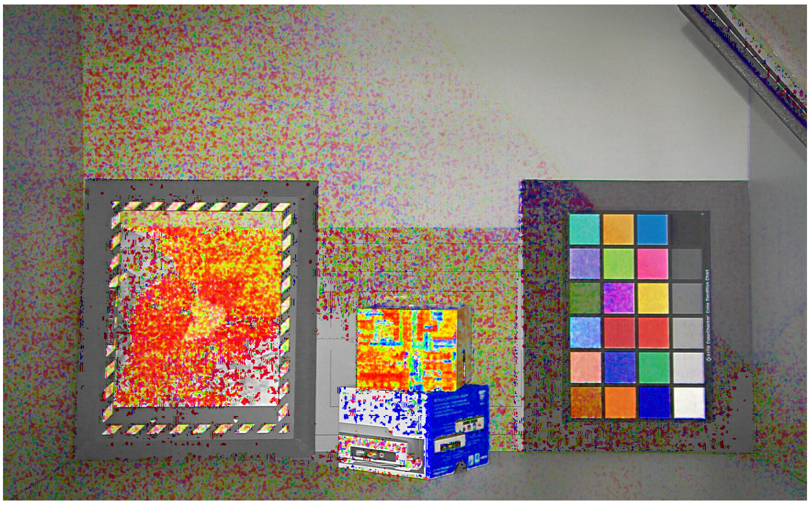

Figure 3.6: Resulting image when the V channel in the color image simply is replaced by the grayscale image based on the NIR and visible light.

3.2.3 Improved HSV Merge Method

To reduce the noise in the dark region, a Gaussian filter was applied to the S channel, see figure 3.7 a). The Gaussian filter was applied as a convolution with a Gaussian kernel and it essentially smears out the image. A weighted region mask, W, was also applied to the S channel, to make the dark region appear less colored and thus less noisy. The weighted region mask is essentially a smooth step function, calculated according to equation (8), where t and s are parameters which determine the shape of the function. The weighted S channel is calculated according to equation (9). The weighted region mask can be seen in figure 3.7 b), with $t = 0.1$ and $s = 0.1$. The resulting S channel, after application of the Gaussian filter and the weighted region mask, can be seen in figure 3.7 c).

$$W = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{\mathrm{V} - t}{s}\right) \tag{8}$$

$$\mathrm{S}_{\text{new}} = W \cdot \mathrm{S}_{\text{filtered}} \tag{9}$$

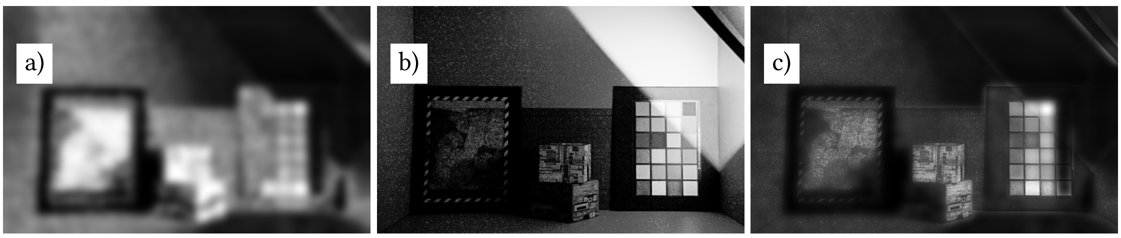

Figure 3.7: The S channel after a Gaussian filter is applied (a), the weighted region mask W (b) and the filtered S channel multiplied with the weighted region mask (c).

The result, with the new S channel and the V channel replaced with the original grayscale image, can be seen in figure 3.8 a) for the parameters $t = 0.1$ and $s = 0.1$, and in 3.8 b) for the parameters

$t = 0.3$ and $s = 0.1$. The result is better and less noisy than the result with the simplest algorithm – compare with figure 3.6. The different parameters give different advantages, for example, decreasing the threshold t makes the colors look stronger but at the same time more noisy.

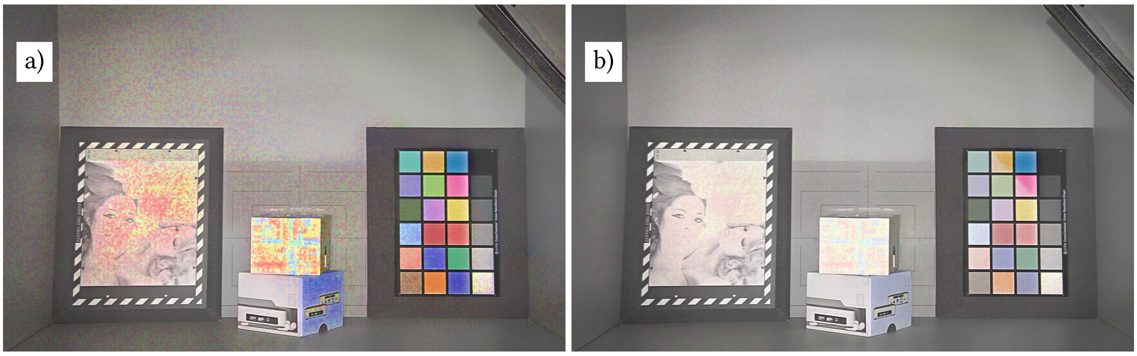

Figure 3.8: Resulting images when a Gaussian filter and a weighted region mask are applied to the S channel. The parameters for the weighted region mask were $t = 0.1$ and $s = 0.1$ (a) and $t = 0.3$ and $s = 0.1$ (b).

3.2.4 YUV Merge Method

To improve the result even more, an algorithm similar to the previous one was used but with YUV space instead of HSV space. The YUV space has the advantage that the intensity channel Y is independent of the color, which is not the case for the V channel in the HSV representation. Furthermore, color channels for YUV were found out to be less noisy than the color channels for HSV – compare figure 3.5 and figure 3.9. Therefore no Gaussian filter had to be applied with the YUV color space. The reason for this is probably that there is a large uncertainty for H and S in the dark regions, as discussed in the background.

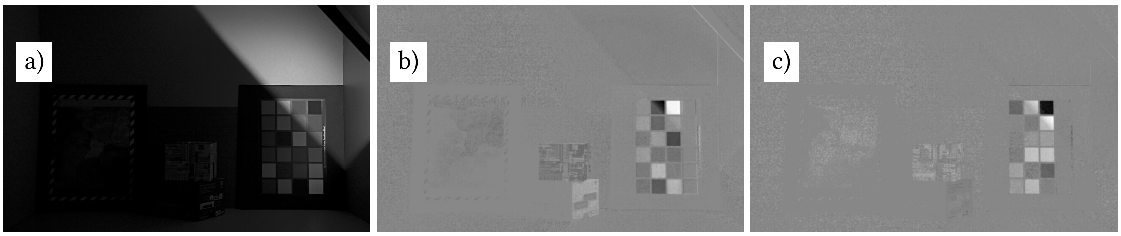

Figure 3.9: The luma channel Y (a) and the chroma channels U (b) and V (c) of the YUV color space.

Like above, a weighted region mask is calculated according to equation (10). But instead of using it to weight the saturation channel S, it is now used to weight the intensity channel Y and the original grayscale image together to a new Y channel. The new Y channel is given in equation (11). The motivation for this approach is to avoid that the intensity of the NIR light to affect the merged image much, if the original color image already is good and bright. The weighted region mask and the new, weighted Y channel are shown in figure 3.10.

$$W = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{\mathrm{Y} - t}{s}\right) \tag{10}$$

$$\mathrm{Y}_{\mathrm{new}} = W \cdot \mathrm{Y} + (1 - W) \cdot \mathrm{NIR} \tag{11}$$

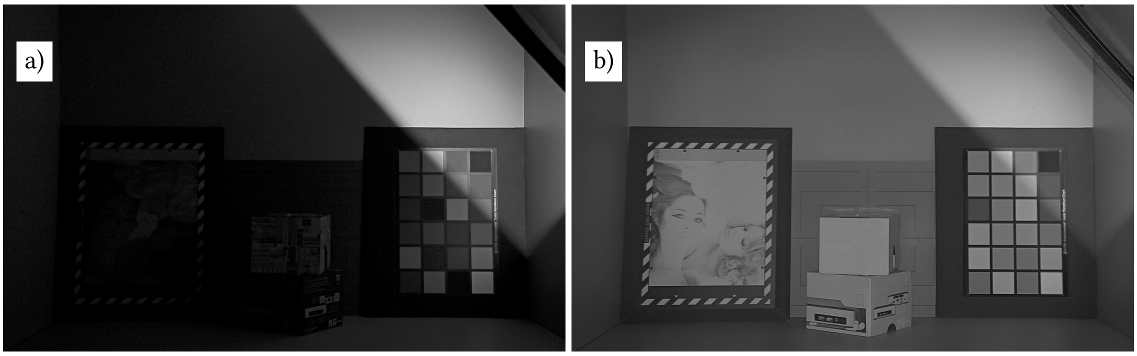

Figure 3.10: Weighted region mask W (a) and the weighted intensity Y_{new} (b).

Combining Y_{new} with the U and V channels gives the image in figure 3.11 a), using the parameters $t = 0.5$ and $s = 0.1$. Compare this with the result from the HSV method, see the images in figure 3.8. The YUV algorithm seems to be producing a noise free image with good color representation. By adjusting the threshold the amount of color in the dark region can be varied, see figure 3.11 b) where the parameters were set to $t = 0.5$ and $s = 1$.

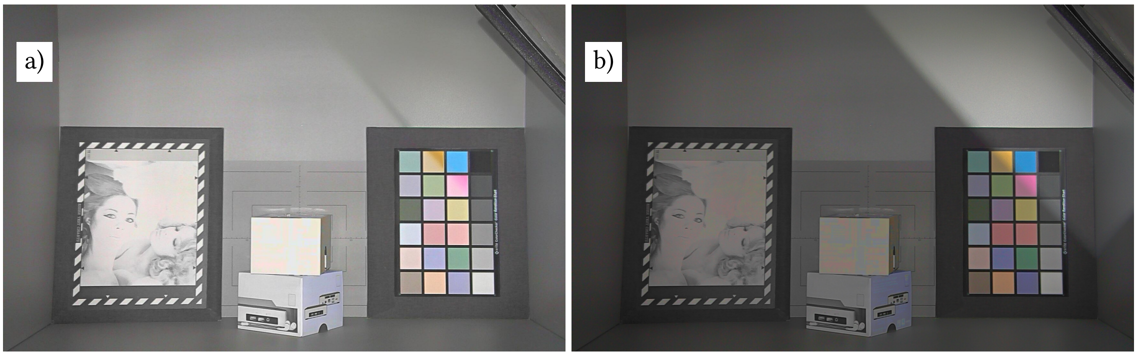

Figure 3.11: Resulting image when the Y channel in the color image is weighted together with the grayscale image according to equation (11). The parameters used are $t = 0.5$ and $s = 0.1$ (a) and $t = 0.5$ and $s = 1$ (b).

The YUV algorithm is the algorithm that worked best and therefore the one chosen to be use with the new camera system. It produces good quality merged images and its easy implementation makes it probably possible to implement it for real-time applications. The weighted region mask makes it possible to take advantage of the information in the grayscale image for dark scenes and at the same time get a good color reproduction in day-time images with highly IR-reflecting objects like trees. To get correct colors in different scenes with different illumination conditions the threshold and smoothness of the weighted region mask need to be adjusted. This is discussed in more detail in the next section.

3.2.5 Choosing Parameters

The YUV merge method has two parameters, t and s, which need to be adjusted to fit different scenes. They determine the shape of the weighted region mask, which in turn determines how the grayscale image and Y channel should be weighted and combined. The impact of s and t on the weighted region mask can be seen in figure 3.12, for some different parameter values.

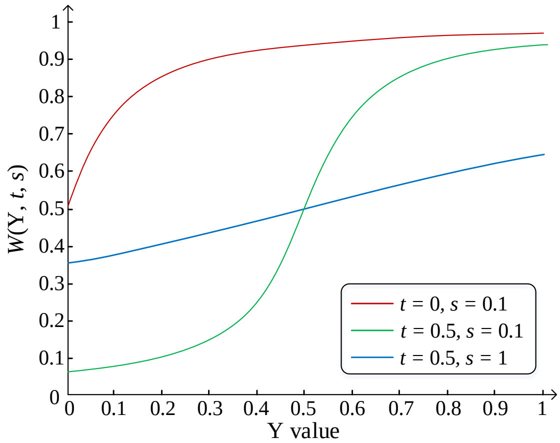

Figure 3.12: The impact of the parameters t and s on the weighted region mask W which is used to merge the grayscale image and the Y channel together.

Choosing parameters is often a balancing act between enhancing details in dark areas on the one hand and obtaining clear colors on the other hand. To illustrate this, the images in figure 3.4 are merged together with different parameters, and the result can be seen in figure 3.13.

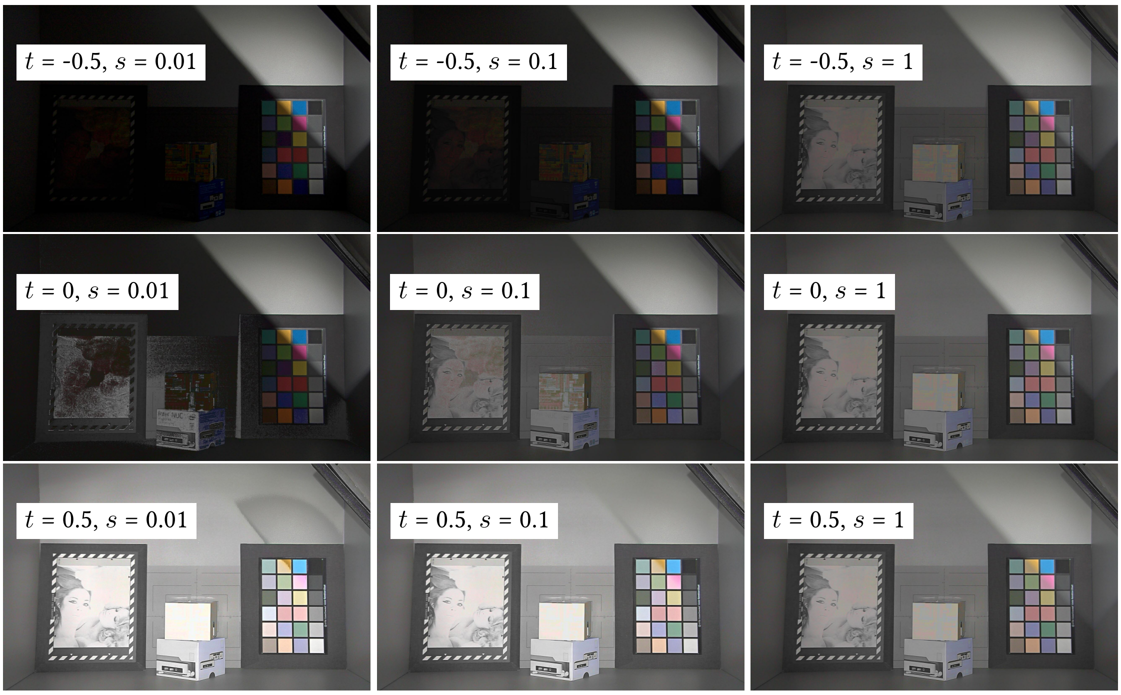

Figure 3.13: The result when the images in figure 3.4 are merged together with the YUV algorithm, for different values of the parameters t and s.

It is clear that different parameter values have different advantages, and for this scene the parameters $t = 0$ and $s = 1$, $t = 0.5$ and $s = 1$ or $t = 0.5$ and $s = 0.1$ seem to give a good result. But the optimal parameters differ for different scenes and different illuminations, and to illustrate this four different scenes with different illumination conditions are shown in the following.

In figure 3.14-3.17, the a) and b) images show the original color and grayscale images. Using the parameters $t = 0.5$ and $s = 0.1$ meant not ideal but good results always were obtained. Therefore, these parameters are chosen as default values and the result using these parameters is shown in image c). Finally, image d) displays a merge image using the parameters that in our opinion yielded the best results.

In figure 3.14 the result of images captured of the same lab scene but this time with poor and even visible illumination in addition to the good IR illumination is given. The merged result combines all information from the color and the grayscale image, giving both good contrast and color information. The different parameter values give similar results, but the optimal values $t = 0$ and $s = 0.1$ give the advantage that the colors are more clear.

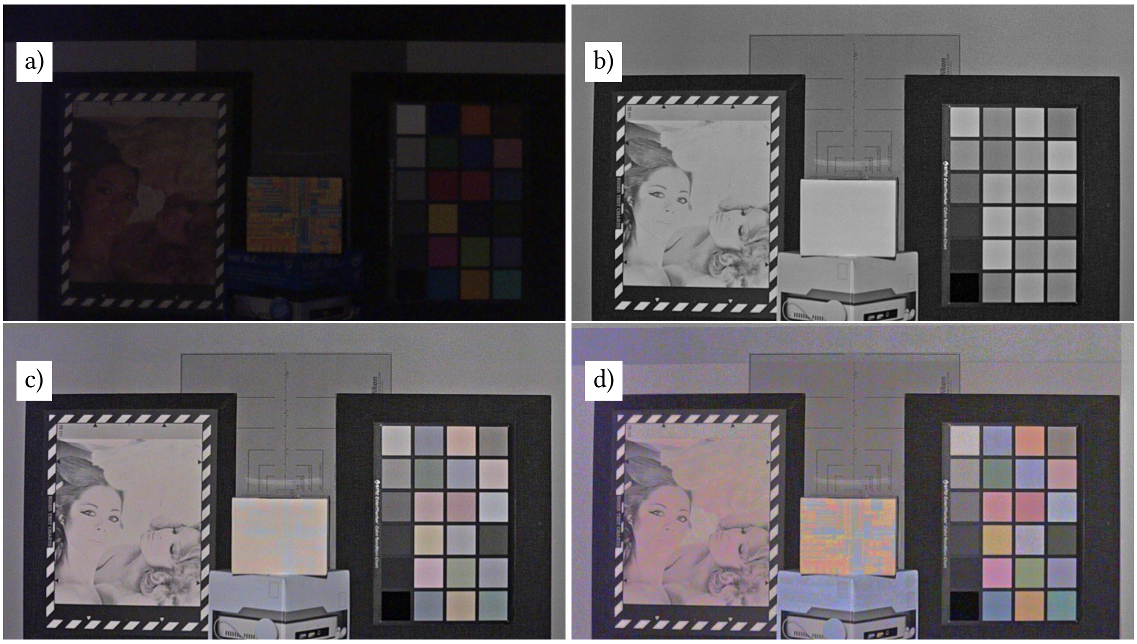

Figure 3.14: A lowlight lab scene. The color image (a), the grayscale image (b), the merged result with the default parameters (c) and the ideal parameters $t = 0$ and $s = 0.1$ (d).

The result of another lab scene can be seen in figure 3.15. Also here the advantage with the ideal parameters $t = -0.3$ and $s = 1$ is that the colors are clearer. But it is harder to determine which parameters that actually are optimal, since some details in the background disappear. Note that the images can be merged to a nice result even if half of the color image is completely shadowed.

Figure 3.15: A partly illuminated indoor scene of a human. The color image (a), the grayscale image (b), the merged result with the default parameters (c) and the ideal parameters $t = -0.3$ and $s = 1$ (d).

An outdoor scene in daylight of a parking lot is given in figure 3.16. Here the main problem is that the grass and the trees reflect a lot of IR light and therefore appear very bright in the grayscale image. To get a natural color it must be avoided to take too much information from the grayscale image, compare (c) and (d). Thus the ideal parameter values, $t = 0$ and $s = 0.1$, are chosen so that the result is similar to the original color image.

Figure 3.16: A day-light outdoor scene of a parking lot including grass and trees. The color image (a), the grayscale image (b), the merged result with the default parameters (c) and the ideal parameters $t = 0$ and $s = 0.1$ (d).

In figure 3.17 the result of an outdoor scene captured during nighttime is given. The main advantage with the optimal parameter values, $t = 0$ and $s = 1$, is that details on the ground and in the sky are more clear. Note that the black jacket reflects IR well meaning that it appears bright on the grayscale image. This, in turn, means that the jacket also is bright in the merged image and even gets a blue shade. This shows a case where one has to be careful using this method.

Figure 3.17: A dark outdoor scene of a human. The color image (a), the grayscale image (b), the merged result with the default parameters (c) and the ideal parameters $t = 0$ and $s = 1$ (d).

3.3 Additional Image Processing

3.3.1 Image Signal Processing

As mentioned before, in digital cameras a lot of image processing is preformed on each image in the camera before the image is displayed to the user or saved. In this project, these kinds of algorithms developed by Axis Communications were used to improve the result. Thus no new algorithms were written, but instead some parameters in the existing algorithms were changed and some features were disabled. For example, the white balancing algorithm had to be adjusted when the dichroic beam-splitter was used since it split too little red light onto the color image sensor. Even more troubling was that the red light varied throughout the image sensor because of the beam-splitter – see results section for details and example of this effect.

Another image signal processing algorithm which was examined during this project was the local tone mapping. This algorithm essentially detects darker areas and makes them brighter, see figure 3.18. Since a lot of testing was done at scenes with poor illumination, this algorithm made the results slightly unpredictable. Also, it might had been better to apply this algorithm on the merged image instead of on the original images. Ultimately, the tone mapping was left on since otherwise no fair comparison could be done with other cameras that automatically had their tone mapping algorithms on.

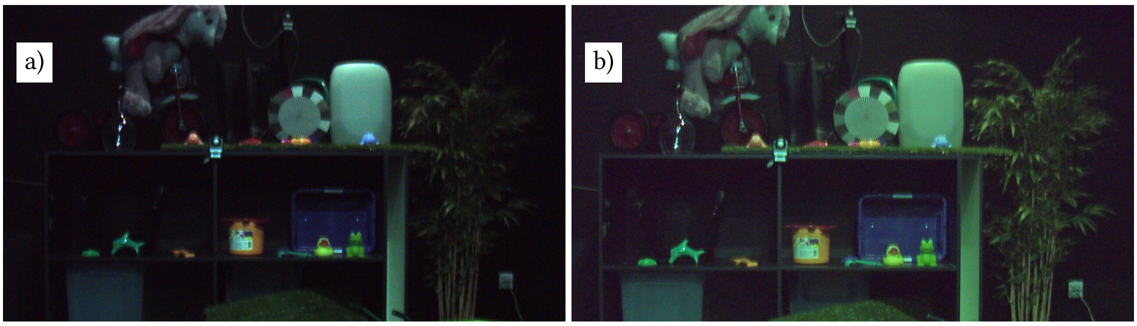

Figure 3.18: Images without (a) and with (b) tone mapping algorithm. The tone mapping makes the image brighter without making it too noisy, at least that was the case in this scene.

3.3.2 Digital Alignment

As discussed in the background section, some sort of digital alignment between the color and grayscale images had to be performed before the images were merged together. The perspective transform was used for this, see the background section for details.

Once the image sensors had been aligned as precisely as possible inside the actual camera, one grayscale and one color image could be captured and used for calibration. The calibration only needs to be done once, at least as long as the image sensors are not touched. In practice, the calibration was done by letting a user click four corresponding points in the two images. Then, the mathematical image transformation to make the corresponding points coincide is calculated and saved.

3.3.3 Performance Issues

The primary goal of this project was to show a proof of concept and therefore not much time was spent on optimizing the algorithms to run fast. The merge algorithm ran offline on a computer instead of in real-time embedded on the camera. Therefore, not much attention was paid to speed but for the future, when the algorithm possibly will be implemented to run in real-time on cameras, it is crucial that this will work. Most of the code was written in Python but when some of the code was implemented in C it became clear that at least the computer could run the algorithm at 30 frames per second and full format. One way to speed up the code would be to replace the tangent function used to get the weighted region mask with a piecewise linear function. After talking to different people at Axis Communications, we feel confident that the developed algorithm would be able to run in real-time embedded on the camera too.

4 Results

4.1 Different Beam-splitters

Choosing and evaluating different beam-splitters has been a major part of this project. Two different kinds were tested; one dichroic thick beam-splitter and one non-dichroic thin beam-splitter and they turned out to have different benefits. First, the level of blur caused by the different beam-splitters is estimated by simulating ray-optics. Then, different images are compared in regard to blur and light sensitivity. The f-number was also changed to see what difference to the level of blur it made.

4.1.1 Sharpness

In order to quantify the differences between different beam-splitters, the ray-tracing optics simulation software Zemax was used. In the simulations, a camera lens similar to the one used in the real setup was used. Simulations were done for the cases when light rays enter the lens at 0^{o} and 45^{o} respectively. In each simulation, all rays would hit the same point on the sensor if an ideal system was used. Because of limitations of the lens and more important the beam-splitter, this was not the case and the rays hit the sensor in a smeared out point whose distribution is called the point spread function. The more point-like this function is, the less blurry will the final image appear.

The results are displayed in table 1 and figure 4.1. RMS@0deg and RMS@45deg denote to the root-mean-square value of the point spread function, assuming that the light enters the camera lens at 0^{o} and 45^{o} respectively. In case a), the beam-splitter was not tilted so this case can be seen as a reference of how well the system possibly could be, since the camera lens actually is designed to have some amount of glass in between it and the image sensor. In case b), the first setup with the 0.2 mm thick beam-splitter was simulated which gave an RMS point spread function of about 8 µm. In order to decrease this value two parameters were changed. First, the f-number of the lens was increased four times in simulation c). Then, the beam-splitter instead was made 40 times thinner in d).

Table 1: The results from four Zemax simulations.

Case	a)	b)	c)	d)
Thickness [µm]	200	200	200	5
Beam-splitter angle [°]	0	45	45	45
f-number	f/1.4	f/1.4	f/5.6	f/1.4
RMS@0deg [µm]	1.0	8.3	1.7	1.2
RMS@45deg [µm]	3.1	8.0	2.6	3.2

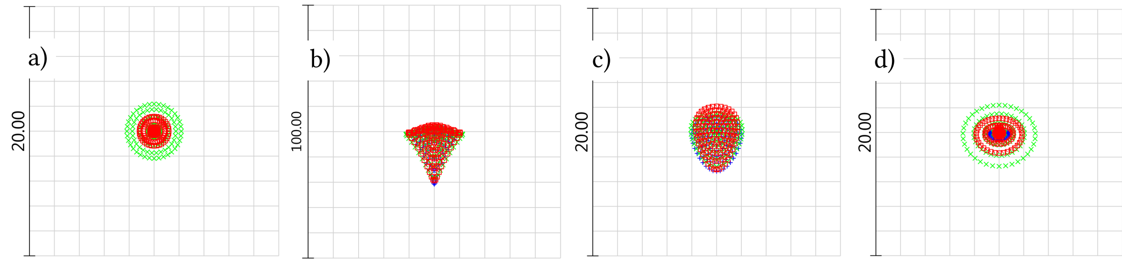

Figure 4.1: Zemax simulations of the point spread function with four different beam-splitter setups assuming that the light enters the lens at 0°. The scales are in µm. The colors represent red, green and blue light. For explanation of the different cases, see table 1.

There are two reasons why the RMS of the point spread function decreases as the aperture decreases. First, the spherical aberration from the camera lens will decrease as the aperture decreases, as described in the background section. Second, the distribution of spatial shifts caused by the beam-splitter will be narrower as the aperture decreases. In an ideal case, without spherical aberration from the camera lens and without any spatial shifts caused by the beam-splitter, the RMS of the point spread function would instead increase as the aperture decreases due to the airy disk [1].

To show that the result from the Zemax simulations actually describes noticeable effects to the image quality with different setups, images of the same scene were captured using the different beam-splitters. This is shown in figure 4.2, where it clearly can be seen that the thicker beam-splitter introduces blur that was not present in the other image, which was captured using the thin pellicle beam-splitter.

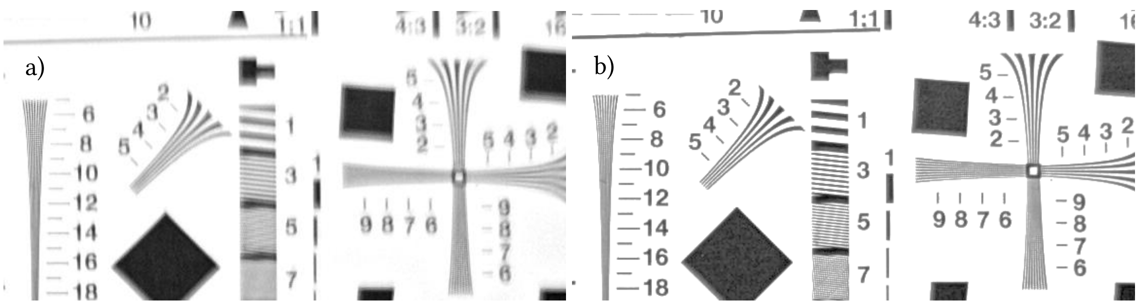

Figure 4.2: Images of a test chart captured using the 0.2 mm thick dichroic beam-splitter (a) and the 2 µm thick pellicle beam-splitter (b), both at 45°.

As could be seen in the Zemax simulations, higher f-numbers partly counter the blur introduced by a thick beam-splitter. This is explained in the background section and verified in figure 4.3 which shows two images captured using the 0.2 mm thick dichroic beam-splitter. Clearly the higher f-number used in b) makes the image look sharper. Keep in mind that the big drawback, especially for security cameras, with using a high f-number is that the camera's light sensitivity is decreased.

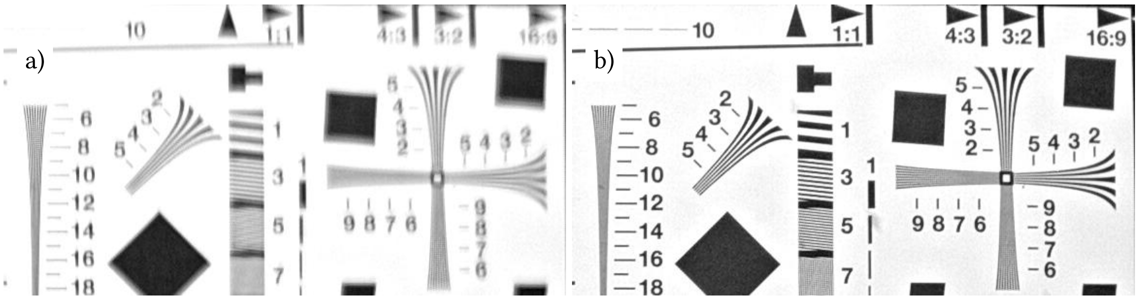

Figure 4.3: Images captured using the 0.2 mm thick beam-splitter and f/1.4 (a) and f/8 (b).

4.1.2 Light Sensitivity

As pointed out in the background section, the pellicle beam-splitter used only reflects approximately 50 % of the visible light when it ideally would reflect 100 %. This means that much of the color information was lost because the pellicle beam-splitter used was not dichroic. This can be seen in figure 4.4, where a) is without any beam-splitter and b) is with the pellicle beam-splitter and thus based on about 50 % less light. The dichroic beam-splitter was unfortunately not designed to fit this setup and therefore some of the red visible light was lost, as discussed previously. This meant that image c), which was captured using the dichroic beam-splitter, is brighter than image b) but slightly darker than image a).

Figure 4.4: Images captured without any beam-splitter (a), with the pellicple beam-splitter (b) and with the dichroic beam-splitter (c).

When the dichroic beam-splitter was used, the captured images appeared too green since too much of the red light was reflected. This is because the cut-off wavelength of the beam-splitter was designed for light hitting it at 0°. The green shade can be seen in figure 4.4 c). To compensate for this, a white balancing algorithm was needed. But even if this is used, the obtained colors will be slightly off, which can be seen in figure 4.5 b) where the right part of the image is more red and the left part is more green. Compare with 4.5 a), which was captured with the pellicle beam-splitter. The color shift appear since light hitting the different regions of the sensor will in average have different angles of incidence on the beam-splitter, which leads to that the cut-off wavelength is different for different parts of the image. Thus a color shift is introduced to the images.

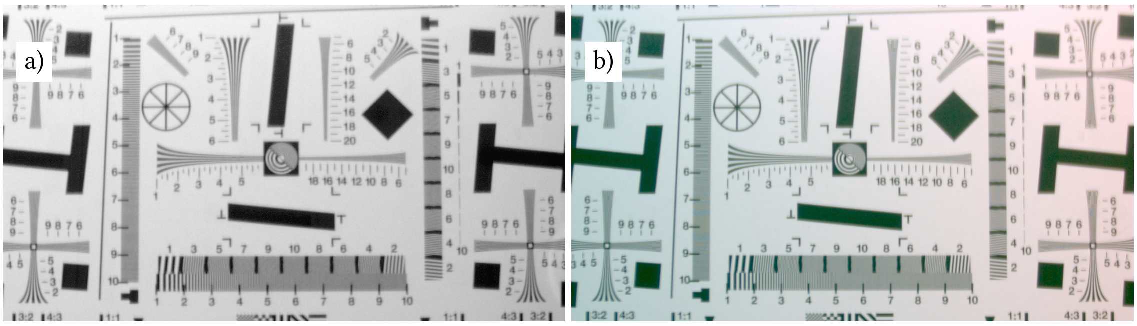

Figure 4.5: Image (a) is captured using the pellicle beam-splitter and gives the correct colors. Image (b) is captured using the dichroic beam-splitter. The color vary due to that the cut-off wavelength is different for different positions of the image.

4.2 Additional Images

The following images were all captured using the setup with the pellicle beam-splitter, since the best results were obtained in this configuration. First, images were captured in a well-controlled lab environment and then in an outdoor environment. For comparison, images of the same scenes but captured with different cameras are also included.

4.2.1 Controlled Lab Scenes

In figure 4.6, images captured of the lab scene that previously has been used are displayed. All of the scene was illuminated by IR LEDs, while only a part of it was illuminated by a flashlight that gave visible light. The merged image contains information that was only present in either the color or the grayscale image. Compare image c), which is the merged image from the developed camera, with image d), which was captured with a camera that uses the same image sensor. This camera has a very wide zoom, giving rise to the Barrel distortion of the image. It is obvious that the concept has some clear benefits when the amount of visible illumination is low. Note that only about 50 % of the visible light could be used by the developed camera due to the beam-splitter, making some of the colored areas in d) look better.

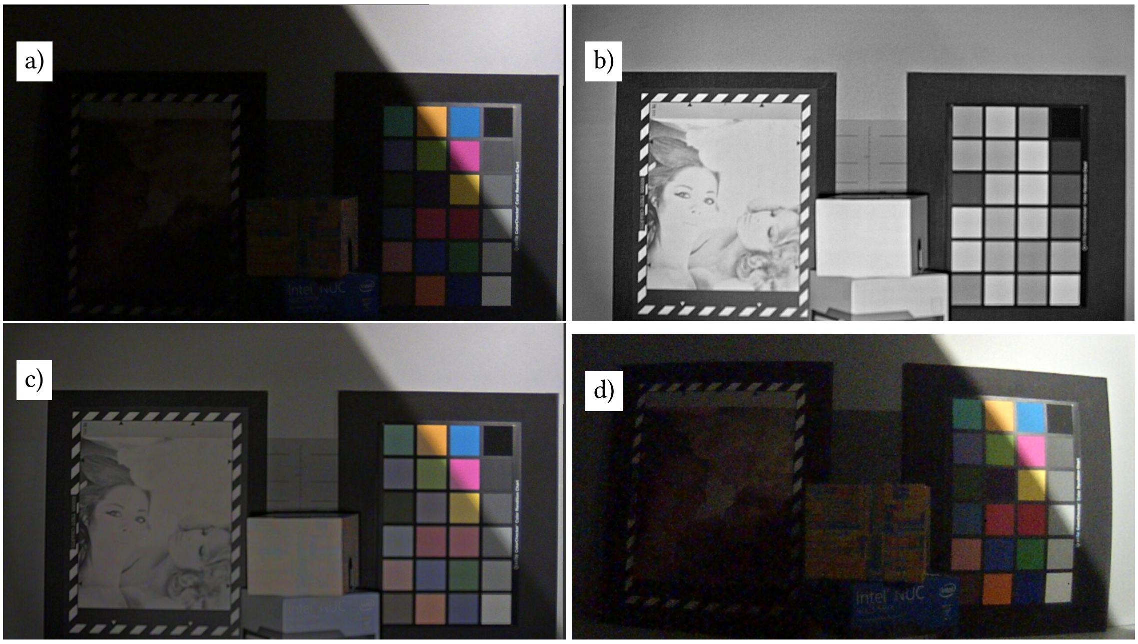

Figure 4.6: A scene with highly varying illumination. Color image (a), grayscale image (b) and merged image (c). Image (d) was captured with a camera from Axis Communications that used the same image sensor as the developed camera.

The camera used to capture the image in figure 4.6 d) did not have the capability to capture both color and grayscale images. Therefore, a comparison with a high-end security camera by Axis Communications using the same scene but slightly different illumination was performed, see figure 4.7. The illumination was chosen such that it was not immediately clear whether the camera should be in day-mode or night-mode, i.e. if it should be capturing color or grayscale images. Exactly this illumination case is where the developed camera setup has an advantage. Image a) and b) were captured by the high-end security camera while c) and d) are the merged results from the developed cameras using different image sensor gains on the color sensor.

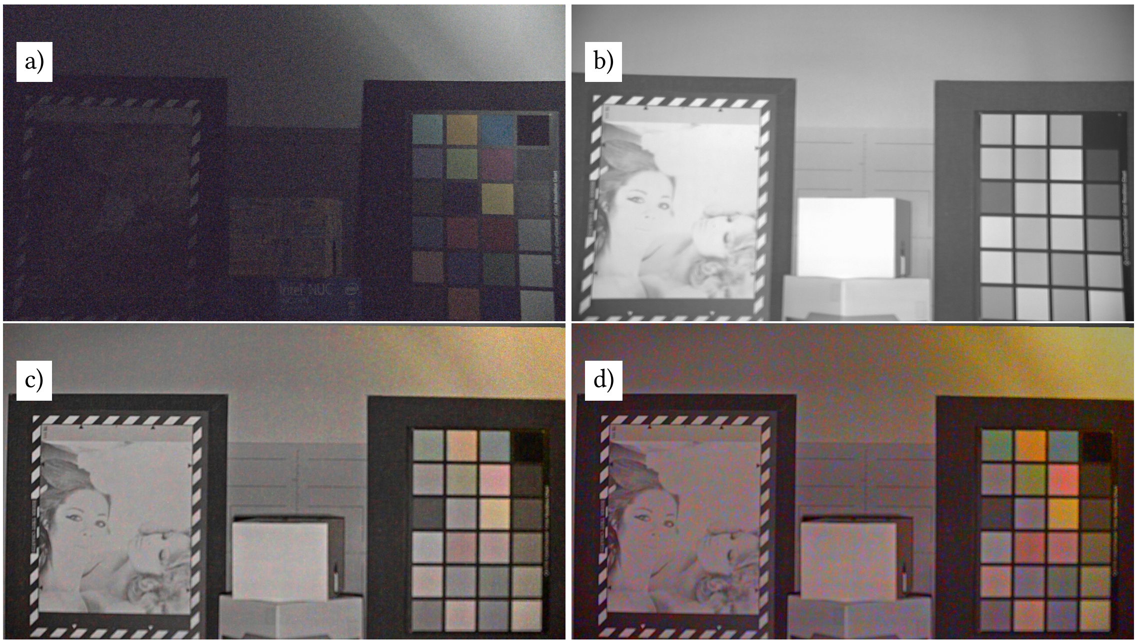

Figure 4.7: A scene with highly varying illumination. Color image (a) and grayscale image (b) were captured using a high-end security camera. Image (c) and (d) are the resulting merged images from the developed camera. For (d), a higher image sensor gain was used which means both clearer colors but also higher noise.

In figure 4.8, a scene more relevant to security purposes was captured. The color image a) has the advantage over the grayscale image b) that texture and color of the shirt can be seen as well as the text and symbols of the signs. On the other hand, the grayscale image is overall much brighter and contains more details such as the knife and the eyes, which cannot be seen in the color image at all. Note that the merged image c) contains all of above mentioned benefits making it the superior choice.

Figure 4.8: Test scene of a human holding a knife. Color image (a), grayscale image (b) and merged image (c).

4.2.2 Outdoor Scenes

To evaluate how well the camera works in outdoor scenes, where the illumination not easily could be adjusted, images were captured outdoors at nighttime. IR LEDs were used to illuminate the scene, but the only visible illumination came from the street lights and the moon. The result can be seen in figure 4.9. Even though the color image is very dark, some features can be extracted from it. Most information in the merged image is taken from the grayscale image which has good contrast but for example the color of the pullover text and the car can possibly be seen too thanks to the color image.

Figure 4.9: Images captured outside at nighttime. Color image (a), grayscale image (b) and merged image (c).

5 Discussion and Conclusion

5.1 Camera Lens and Image Sensor

The camera lens used was IR-corrected, but it still had a slightly different level of zoom for the color and the grayscale image. This was one of the reasons why the digital alignment was needed. Since the visible and NIR light ideally is focused onto different sensors, potentially at different distances from the camera lens in the multi-sensor setup, it is actually not necessary to use an IR-corrected lens like it is in most other security cameras. This opens up opportunities to use a larger range of camera lenses, possibly decreasing cost and enhancing performance.

The choice of image sensors was in this project only determined by the size of the sensors' PCB board. The resources to design a new PCB board were insufficient and only one kind of board was small enough to fit the tight setup. The corresponding image sensor did not have the best light-sensitivity which became clear when the developed camera's performance was compared with other cameras' performances. It should, however, be possible to design a smaller PCB board for some of the high-end image sensors to fit this setup, especially if this camera design would go into mass production. Another aspect worth mentioning is that the same type of image sensor was used to capture both visible and NIR light. In an improved implementation, the sensor capturing NIR light could possibly be optimized to do just this; capturing NIR light instead of visible light.

5.2 Beam-splitter

Much of this report has been dedicated to evaluating beam-splitters. It is clear that it is not possible to use the 0.2 mm thick beam-splitter since it causes too much blur to the transmitted image. Looking at the Zemax simulations, the RMS value of the point spread function was about 8 µm compared to 1 µm without any beam-splitter. It is desirable for the RMS value to at least not exceed the pixel pitch, which in relevant image sensors today is about 2-3 µm. This means that the beam-splitter is not allowed to be thicker than about 50 µm according to the Zemax simulations. Only two beam-splitters have been tested, one that was 200 µm thick and another that was 2 µm thick. It was very clear that the thicker beam-splitter caused too much blur, while the thin one managed to produce good-quality images.

One desirable feature of the beam-splitter used in the setup is that it should be dichroic and hence split the visible light onto one sensor and the NIR light onto the other sensor. Only the thick beam-splitter that was tested had this capability and it seemed to work reasonably well, except for the blur introduced because of above mentioned reason. Another problem was that the cut-off frequency for the beam-splitter was slightly wrong, meaning that some of the red light that was supposed to hit the sensor capturing color images instead hit the other image sensor. It would, however, be possible to obtain a beam-splitter with correct cut-off frequency designed for 45° angle of incidence.

The focus of the project was to build a prototype camera that proved the concept as a whole, and therefore not much time was spent on examine if it would be possible to obtain a dichroic pellicle beam-splitter. On the one hand, we could not find any dichroic pellicle beam-splitters to purchase and also think that adding a multi-layer coating could add strain and deformation to the membrane. On the other hand, some sort of coating is already applied to some pellicle beam-splitters and also multi-layer coatings for pellicle beam-splitters have been discussed previously, see for example the Hewlett-Packard patent [22]. Obtaining a dichroic pellicle beam-splitter is perhaps the first and most important upgrade to the current system.

Finally, a thorough evaluation of pellicle beam-splitters' suitability for security cameras would had been needed before this camera would go into mass production. Although the sample we used never broke, the 2 µm thick membrane is supposedly very fragile which possibly could add complications. On the other hand, pellicle beam-splitters are already used in some DSLR cameras for the viewfinder and the autofocus which would suggest that it would be possible to introduce pellicle beam-splitters in security cameras too.

5.3 Merge Algorithm

Overall, the algorithm that merges the color and grayscale images into one superior image is working satisfactory. In contrast to some of the algorithms developed previously, the algorithm developed here is simple and easy to both understand and to implement. Another benefit is that the algorithm is fast and should be able to run in real-time on the camera. The fact that two user-defined parameters are used can be both positive and negative. It is good in the sense that the algorithm can be tuned and optimized to work well in different environments. On the other hand, it would had been great to have a completely general algorithm where the parameters would not have to be changed as the illumination, for example, changed. Our conclusion is, however, that the default parameter values work well, at least for the scenes we have worked with.

In the future it should be possible to tune the parameters automatically. For instance, the current exposure time and gain for the color image sensor could be used to draw conclusions on the illumination conditions. With that information it should be possible to make an assessment of how the color and grayscale image should be weighted together. Generally, one wants to extract much information from the grayscale image in dark scenes and less in brighter scenes to avoid false colors.

One issue with the algorithm and this concept in general is that certain objects reflect NIR light much more than visible light. This means that humans think of the object to be relatively dark while a grayscale image based on NIR light will display the object as much brighter. This means that the developed algorithm typically will apply the correct color shade to such objects but reconstruct the brightness wrongly, which sometimes makes the object look unnatural. We believe that this is a fundamentally difficult problem which to some extent only can be compensated for instead of fully removed.

While this might not be a large problem with trees and grass, since their color usually is not that interesting for security applications, it can be a problem with some fabrics which reflect NIR light very well. This can for example be seen in figure 3.17, where the black jacket gets a bright and wrong color. One of the main motivations for this project was to obtain colors also in low light conditions and it has been a clear goal to try to get as correct colors as possible at all times.

5.4 Resulting Images

In order to be able to make a fair comparison between the resulting setup and a security camera, images of the same scene were captured with both this setup and with a camera that uses the same image sensor. The result can be seen in figure 4.6. On the one hand, the security camera, image d), captures more colors because the pellicle beam-splitter splits of about half of the visible light. But on the other hand, the contrast is much better in the merged image, image c), because of the grayscale image. Overall, we think that the merged image is to prefer even though some color is lost.

Images of the same scene but with a slightly lower illumination were captured both with a high-end security camera, see figure 4.7 a) and b), and with the developed setup using two different values of the sensor gain, see figure 4.7 c) and d). With both values of gain, some color information could be extracted from the color image and added to the merged image. Therefore, the results from the developed camera were better than the other camera's result which were either noisy color images or colorless but clear grayscale images. The color image captured with the high-end camera has more color information than what was captured with the multi-sensor camera, but once again this is, at least partly, explained by the fact that the pellicle beam-splitter splits half of the visible light to the wrong sensor. But since the painting to the left in the test scene hardly can be seen at all because it is too dark, the merged result is preferable. Note that the image sensor used in the high-end security camera is better than the ones used in the prototype camera, and thus this comparison is not completely fair.

In more security camera relevant scenes, like those in figure 4.8 and 4.9, the concept works well too. In many scenes, no big difference can be seen between the images captured with the developed setup and with an ordinary security camera, since it will be clear whether the day-mode or the night-mode gives the best image. But often at least some colors can be extracted even during night time and thus there clearly is potential for this concept. According to theory, the concept could potentially have benefits also in hazy conditions. During this project we were never able to capture any images where this effect could be seen though. We did try to see the same effect in misty weather conditions but no improvement could be seen, possibly because the particles had such size so that the difference in Rayleigh scattering for visible and NIR light was too small.

5.5 Competitive Solutions

One competing solution that should be mentioned is the idea of just mounting two cameras beside each other. One camera would capture color images based on visible light and the other would capture grayscale images based on NIR light. The two images would then be digitally aligned and merged just like they are in the multi-sensor camera. This approach is much simpler to implement and one does not need to worry about the beam-splitter. On the other hand, the multi-sensor solution is probably cheaper and also the parallax problems that the solutions with two separate cameras automatically will have are avoided.

Another solution is to use an image sensor with another color array filter instead of the conventional Bayer filter. For instance, half of the pixel filters that transmit green light in a Bayer filter could be replaced by pixel filters that transmit only NIR light. This solution is very clean and it does not require any beam-splitter or second image sensor. One problem, however, is that the camera lens would need to be able to focus light from a very large spectral range onto the same focus plane to avoid blur. This may be both hard and expensive. Another problem is that the intensity of NIR light in some situations will be significantly higher than the intensity of visible light. This can be a problem because the optimal exposure time would be very different for the different pixels.

A third alternative solution would be to use another lens system where the beam-splitter is positioned in between two internal lenses of the camera lens. It should be placed in a spot where the light is approximately collimated, which in turn would imply that the blur caused to the transmitted image by a thicker beam-splitter would be much smaller. Therefore, one would not need to use a pellicle beam-splitter but could instead use a dichroic glass beam-splitter and would avoid many of the problems with the current setup. This approach may, however, be hard to put in mass-production and it would also be more expensive since a customized lens system would need to be developed and manufactured.

One final approach that should be mentioned is to develop a concept inspired by 3CCD and 3MOS cameras. These cameras have a prism after the lens to split the red, green and blue light onto three different image sensors. In a security camera this prism would, just like the beam-splitter, split the visible and NIR light onto two separate image sensors. The advantage of using a prism over a beam-splitter is that the light can be split up based on its wavelength and at the same time the prism does not introduce any significant blur. This approach would require a special lens and the prism would probably make the solution more expensive.

5.6 Considerations Going Forward

We believe that the concept investigated here is promising and that it has advantages over some of the current security camera designs. In certain situations it has been shown that this multi-sensor camera outperforms today's high-end security cameras. This was the case with the current setup and it would certainly be the case in an upgraded setup with better alignment, a solution for the beam-splitter problem and better image sensors. There are, however, some questions that were not investigated in great detail and which would need to be answered before this concept is brought to the next step. First, it needs to be determined if a dichroic pellicle beam-splitter with desirable characteristics is possible to manufacture. Then, a solution to the alignment problem need to be found. Today, there are systems to align one sensor on mass production scale and therefore it is probably possible to build an alignment system which can align two sensors and possible the beam-splitter all at once.

Another aspect worth considering is the business aspect. While we feel confident that this solution provides a better result in certain imaging situations, it needs to be determined if the advantage is sufficiently large and if the relevant situations occur sufficiently often. After all, most of the time security cameras capture images in situations where it is absolutely clear if the cameras should be in day-mode or night-mode and thus the improvement the developed camera could bring may be limited. On the other hand, there are a few more situations where this solution might be advantageous such as the situations where part of an image ideally would be in day-mode and another part in night-mode. Also, there may be additional benefits with having two sensors that were not investigated here. For instance, the different image sensors could use different exposure time so that the camera achieves high dynamic range (HDR) with synced images. This could potentially remove some of the problems with HDR today where moving objects looks wrong because the same image sensor captures both the images with short and long exposure time.

As far as cost is concerned, the only parts that need to be added to a conventional camera in the multi-sensor solution are one image sensor and the beam-splitter. The image sensor is relatively cheap compared to the rest of a security camera, but how expensive the beam-splitter would be is unknown. However, some DSLR cameras already use pellicle beam-splitters. This suggests that these elements should be viable to implement in security cameras too. It may also be possible to remove some of the parts that sometimes are used in other security cameras. For example, the light sensor which determines whether the camera should be in day-mode or night-mode may be unnecessary when two image sensors are used, which contributes to a lower cost.

Bibliography

[1] B. Saleh and M. Teich. *Fundamentals of Photonics.* John Wiley & Sons, 2007. ISBN: 978-0-471-35832-9.

[2] B.E. Bayer. *Color imaging array.* US Patent 3,971,065. 1976.

[3] R. Turchetta, K. Spring and M. Davidson. URL: http://olympus.magnet.fsu.edu/primer/digitalimaging/cmosimagesensors.html (visited on 12/01/2015).

[4] E. R. Fossum. "Active pixel sensors: are CCDs dinosaurs?" In: *Proc. SPIE* 1900 (1993), pp. 2–14.

[5] Thorlabs. URL: https://www.thorlabs.de/newgrouppage9.cfm?objectgroup_id=3313 (visited on 12/01/2015).

[6] Thorlabs. URL: https://www.thorlabs.de/NewGroupPage9.cfm?ObjectGroup_ID=898 (visited on 12/01/2015).

[7] H. Telle, A. Ureña and R. Donovan. *Laser Chemistry: Spectroscopy, Dynamics and Applications.* John Wiley Sons, 2007. ISBN: 9780470059401.

[8] R. Szeliski. *Computer Vision: Algorithms and Applications.* Springer, 2010.

[9] M. Horvath. URL: https://commons.wikimedia.org/wiki/File:RGBCube_b.svg (visited on 12/22/2015).

[10] URL: http://www.compression.ru/download/articles/color.space/ch03.pdf (visited on 12/01/2015).

[11] A. Hanbury and J. Serra. "A 3D-polar Coordinate Colour Representation Suitable for Image Analysis". In: *Computer Vision and Image Understanding* (2002).

[12] SharkD. URL: https://en.wikipedia.org/wiki/HSL_and_HSV#/media/File:HSV_color_solid_cone_chroma_gray.png (visited on 01/09/2016).

[13] R.I. Olsen, J. Gates, and D.L. Sato. *Digital cameras with direct luminance and chrominance detection.* US Patent 7,964,835. 2011.

[14] S. Eugster. URL: https://commons.wikimedia.org/wiki/File:YUV_UV_plane_Y0.5_100_percent.png (visited on 12/22/2015).

[15] X. Zhang, T. Sim and X. Miao. "Enhancing Photographs with Near Infrared Images". In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (2008), pp. 1–8.

[16] H. Honda, R. Timofte and L. Van Gool. "Make My Day - High-Fidelity Color Denoising with Near-Infrared". In: *CVPR Workshops. IEEE* (2015).

[17] M. Brown and S. Süsstrunk. "Multi-spectral SIFT for scene category recognition". In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (2011), pp. 177–184.

[18] C. Fredembach, N. Barbuscia and S. Süsstrunk. "Combining visible and near-infrared images for realistic skin smoothing". In: *Color and Imaging Conference, 17th Color and Imaging Conference Final Program and Proceedings* (2009), pp. 242–247.

[19] N. Salamati, A. Germain and S. Süsstrunk. "Removing Shadows from Images Using Color and Near-Infrared". In: *Image Processing (ICIP), 2011 18th IEEE International Conference on* (2011), pp. 1713–1716.

[20] C. Feng, S. Zhuo, X. Zhang, L. Shen and S. Süsstrunk. "Near-Infrared Guided Color Image Dehazing". In: *Image Processing (ICIP), 2013 20th IEEE International Conference on* (2013), pp. 2363–2367.

[21] S. Svanberg. *Atomic and Molecular Spectroscopy.* Springer, 2004. ISBN: 3-540-20382-6.

[22] J.S. Yeo, S.V. Mathai, and M.R.T. Tan. *Fabrication of thin pellicle beam splitters.* US Patent 8,711,484. 2014.