

HYPER SPECTRAL IMAGING

F. SIGERNES

AGF-331: Remote Sensing and Spectroscopy – lecture notes

CONTENS

1	BASIC SPECTROSCOPY	
	1.1 Light as waves	1
	1.2 Interference	2
	1.3 Diffraction	4
	1.4 The GRISM	9
2	SPECTRAL DESIGNS	
	2.1 The Ebert-Fastie spectrometers	11
	2.2 The Czerny-Turner configuration	13
	2.3 The FICS concave grating spectrograph	
	2.4 The GRISM spectrograph	14
	2.5 The Auroral Spectrophotometer number 3 (SP3)	15
3	SYSTEM OPTICS	
	3.1 Spectrometer optical diagram	17
	3.2 Front optics	18
	3.3 f/value of a spectrometer	
	3.4 Magnification of the slits	19
	3.5 Bandpass and resolution	
4	THROUGHPUT AND ETENDUE	
	4.1 Definitions	22
	4.2 Etendue	
	4.3 Flux	24
	4.4 Throughput	
	4.5 Stray light	
5	IMAGING SPECTROSCOPY	
	5.1 Short background	27
	5.2 A low cost hyperspectral imager	
	5.3 Hyperspectral image formation and spatial resolution	28
	5.4. Example applications	31
	5.5. What's next?	36
6	CALIBRATION	
	6.1 Spectral narrow field of view calibration	37
	6.2 Bandpass filter instrument calibration	45
	6.3 Color camera calibration	46
7	ATTITUDE CORRECTIONS	
	7.1 Introduction	54
	7.2 Least squares adjustment	
	7.3 Kalman filtering	57
	7.4 Airborne attitude estimation	59

8 FUTURE**References**

Appendix A: Basic equations for an Ebert-Fastie spectrometer

Appendix B: Basic calculations for a 400 – 900 nm GRISM spectrograph

Appendix C: Spatial resolution calculations with different front optics

1. BASIC SPECTROSCOPY

In order to explain how a spectrometer works, it is necessary to refresh some basic concepts of physics like interference and diffraction.^{1,2} The key optical element of a spectrometer is the diffraction grating. In this chapter we will derive the grating equation, resolving power, angular dispersion, and grating efficiency.

1.1 Light as waves

First, let us define that light is tiny packets called photons with wave like properties. Fig.1.1 shows a sinusoidal wave that moves with velocity v along the x-axis. The amplitude of the wave can then be defined as simply

$$E(x,t) = E_0 \sin(kx \pm \omega t), \quad (1.1)$$

where E_0 is the maximum amplitude of the wave. The wave repeats itself periodically in time t by $T = \lambda/v$, where λ is the wavelength and v is equal to the speed of light, c . $k = 2\pi/\lambda$ is called the propagation number or wave number, and angular frequency is $\omega = 2\pi/T$.

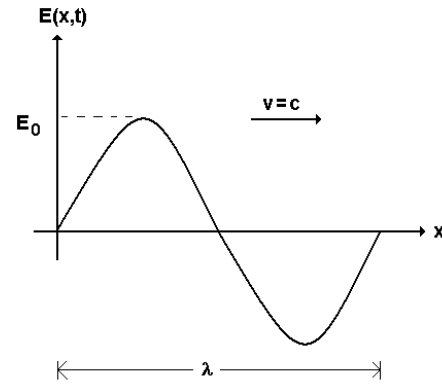


Figure 1.1. Sinusoid wave.

A more convenient way of representing a wave is to use complex mathematics.² A wave can be described as the real part of a complex number (see Fig. 1.2)

$$z = x + iy \quad (1.2)$$

$$z = E_0(\cos \phi + i \sin \phi).$$

Furthermore, Euler's formula states that

$$e^{i\phi} = \cos \phi + i \sin \phi, \quad (1.3)$$

where is $\phi = kx \pm \omega t$. The real part of z is then

$$\text{Re } z = E = E_0 \cos \phi. \quad (1.4)$$

This is a very useful form when dealing with waves due to rules of complex mathematics. In 3D we may express the wave as

$$\mathbf{E}(\mathbf{r}, t) = E_0 e^{i\phi} \quad (1.5)$$

where \mathbf{r} is the position and the phase is defined as

$$\phi = \mathbf{k} \cdot \mathbf{r} - \omega t + \xi. \quad (1.6)$$

ξ is the initial phase of the wave.

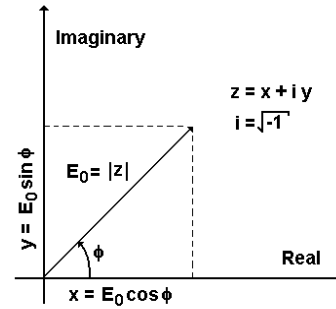


Figure 1.2. Complex number.

1.2 Interference

1.2.1 Two wave sources

Let us consider the simple case when just two waves S_1 and S_2 act together. According to equation (1.5) and (1.6), the waves may be expressed as

$$\begin{aligned} \mathbf{E}_1 &= \mathbf{E}_{01} e^{i\phi_1} \\ \mathbf{E}_2 &= \mathbf{E}_{02} e^{i\phi_2} \end{aligned} \quad (1.7)$$

where the phases are

$$\begin{aligned} \phi_1 &= \mathbf{k}_1 \cdot \mathbf{r}_1 - \omega t + \xi_1 \\ \phi_2 &= \mathbf{k}_2 \cdot \mathbf{r}_2 - \omega t + \xi_2 \end{aligned} \quad (1.8)$$

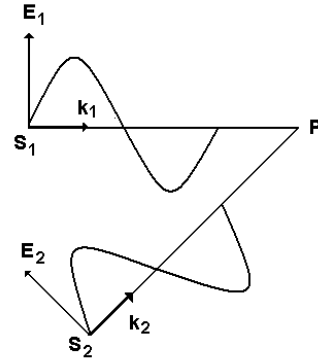


Figure 1.3. Two waves.

At point P the net result must be the sum of the vectors $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$. The intensity is then

$$\begin{aligned} I &\approx \mathbf{E} \cdot \mathbf{E}^* = (\mathbf{E}_{01} e^{i\phi_1} + \mathbf{E}_{02} e^{i\phi_2}) \cdot (\mathbf{E}_{01} e^{-i\phi_1} + \mathbf{E}_{02} e^{-i\phi_2}) \\ &= \mathbf{E}_{01}^2 + \mathbf{E}_{02}^2 + 2\mathbf{E}_{01} \cdot \mathbf{E}_{02} (e^{i(\phi_1 - \phi_2)} + e^{-i(\phi_1 - \phi_2)}) \\ &= \mathbf{E}_{01}^2 + \mathbf{E}_{02}^2 + 2\mathbf{E}_{01} \cdot \mathbf{E}_{02} \cos(\phi_1 - \phi_2) \\ &= \mathbf{E}_{01}^2 + \mathbf{E}_{02}^2 + 2\mathbf{E}_{01} \cdot \mathbf{E}_{02} \cos \sigma \end{aligned} \quad (1.9)$$

The phase difference $\sigma = (\mathbf{k}_1 \cdot \mathbf{r}_1 - \mathbf{k}_2 \cdot \mathbf{r}_2) + (\xi_1 - \xi_2)$ is a sum of two terms. The first term is a result of the path difference between the two waves and the second due to the initial phase difference. Constructive and destructive interference occurs when $\cos \sigma$ equals +1 or -1, respectively.

$$\sigma = \begin{cases} 2n\pi & \text{Constructive interference} \\ (2n+1)\pi & \text{Destructive interference} \end{cases} \quad (1.10)$$

In equation (1.10), n is a positive or negative integer. Note that in the above we assumed that the waves were linearly polarized plane waves.

1.2.2 Several wave sources

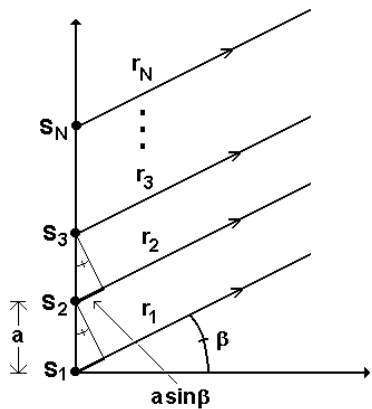


Figure 1.4. N number waves.

Next we consider N number of waves that are emitted by S_N coherent and monochromatic sources. Each source is separated by a distance of a .

$$\begin{aligned} \mathbf{E}_1 &= \mathbf{E}_{01} e^{i(\mathbf{k} \cdot \mathbf{r}_1 - \omega t)} \\ \mathbf{E}_2 &= \mathbf{E}_{01} e^{i(\mathbf{k} \cdot \mathbf{r}_2 - \omega t)} \\ \mathbf{E}_3 &= \mathbf{E}_{01} e^{i(\mathbf{k} \cdot \mathbf{r}_3 - \omega t)} \\ &\vdots \\ \mathbf{E}_N &= \mathbf{E}_{01} e^{i(\mathbf{k} \cdot \mathbf{r}_N - \omega t)} \end{aligned} \quad (1.11)$$

Coherent means that the phase difference between the waves are constant. Wave numbers and amplitudes are the same for all waves considered.

The phase difference between the waves is due to the path difference

$$\sigma = \frac{2\pi}{\lambda} a \sin \beta . \quad (1.12)$$

In Fig. 1.4 we assume that the distance to point P is much larger than a . It is common to use a method called the rotating vector sum method to obtain the total vector \mathbf{E} at point P .

In Fig. 1.5 O is the centre of circle with a radius of ρ . It is circular since σ and $E_0 = E_{01}$ are the same for each source. E is then found as

$$\begin{aligned} E &= CP = 2QP \\ &= 2\rho \sin\left(\frac{1}{2} N\sigma\right). \end{aligned} \quad (1.13)$$

From the triangle QCR we get

$$E_{01} = 2\rho \sin\left(\frac{1}{2} \sigma\right). \quad (1.14)$$

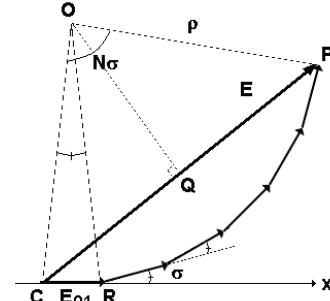


Figure 1.5. Rotating vector sum.¹

Eliminating ρ from equations (1.13) and (1.14) by division we finally obtain

$$E = E_{01} \frac{\sin\left(\frac{1}{2} N\sigma\right)}{\sin\left(\frac{1}{2} \sigma\right)} \quad (1.15)$$

The intensity is proportional to E^2

$$I = I_0 \left(\frac{\sin\left(\frac{1}{2} N\sigma\right)}{\sin\left(\frac{1}{2} \sigma\right)} \right)^2 = I_0 \left[\frac{\sin(N\pi a \sin \beta / \lambda)}{\sin(\pi a \sin \beta / \lambda)} \right]^2. \quad (1.16)$$

Fig. 1.6 shows a plot of equation (1.16) as a function of wavelength λ and angle β . The functions have a maxima equal to $N^2 I_0$ when $\sigma = 2n\pi$. Or, in other words

$$n\lambda = a \sin \beta . \quad (1.17)$$

Zero intensity occurs when $(1/2)N\delta = n'\pi$, where n' is any integer except $n' \in [0, N, 2N, \dots]$. There are $N-2$ minima between maxima. As a consequence, $\Delta\sigma = 2\pi / N$ corresponds to the phase change from peak down to zero intensity, or the half widths of the interference peaks.

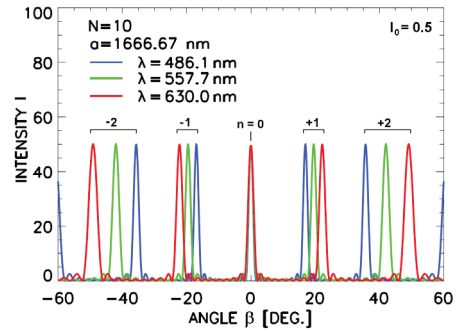


Figure 1.6. Interference as a function of angle β and wavelength λ from $N=10$ sources. $a = 1666.67$ nm. $I_0 = 0.5$.

Note that as the wavelength increases, the profiles shift further away from zero order, $n = 0$. In addition, the separation between wavelengths increases with increased order number, n . As N increases, the width of the peaks decreases. A distance of $a = 1666.67$ nm corresponds to a grating with 600 lines or grooves per mm.

1.3 Diffraction

Interference is the result of individual sources interacting with each other. Diffraction, on the other hand, is noticeable when a wave is distorted by an obstacle that has dimensions comparable to the wavelength of the wave. An example is shown in Fig. 1.7 where ocean waves hit the entrance wall to a port or harbour.

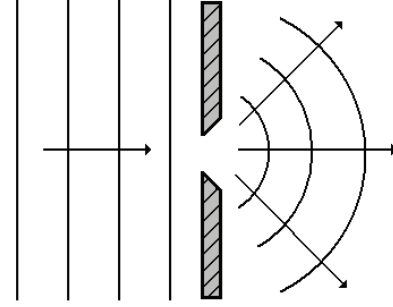


Figure 1.7. Example of diffraction: Ocean waves hitting a port entrance wall.

1.3.1 The single slit

A single slit may be used to diffract light in the same manner as described above. The width of the slit should be in the order of $100\mu\text{m}$ or less to see any effects in the visible range of the spectrum ($400 - 700\text{ nm}$).

Huygens principle states that every point of a wave front can be thought of as the source of secondary wavelets. See red dots in Fig. 1.8. These new waves are defined as the diffracted waves, each with amplitude dE_{01} .

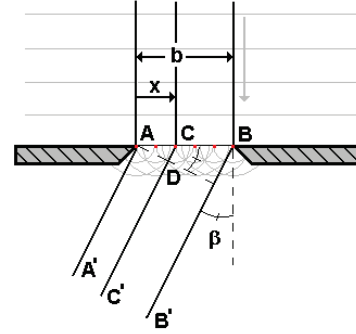


Figure 1.8. The single slit.

The phase difference σ between CC' and AA' is

$$\sigma = \frac{2\pi}{\lambda} CD = \frac{2\pi x \sin \beta}{\lambda}. \quad (1.18)$$

The same method we used in section 3.1 is shown in Fig. 1.9. The length of E becomes

$$E = 2QP = 2\rho \sin\left(\frac{1}{2}\alpha\right). \quad (1.19)$$

When $\beta = 0$ then

$$E_0 = \rho\alpha = \rho\left(\frac{2\pi b \sin \beta}{\lambda}\right), \quad (1.20)$$

since all dE_0 vectors are parallel for observations perpendicular to the slit.

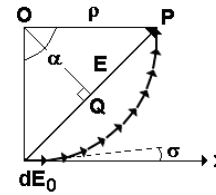


Figure 1.9. Rotating vector sum.¹

From equation (1.19) and (1.20) we get

$$E = E_0 \left[\frac{\sin(\pi b \sin \beta / \lambda)}{\pi b \sin \beta / \lambda} \right], \quad (1.21)$$

and the intensity is proportional to the square of the amplitude

$$I = I_0 \left[\frac{\sin(\pi b \sin \beta / \lambda)}{\pi b \sin \beta / \lambda} \right]^2 = I_0 \left(\frac{\sin u}{u} \right)^2, \quad (1.22)$$

where $u = \pi b \sin \beta / \lambda$.

Fig.1.10 shows the diffraction pattern as a function of wavelength λ and angle β . Zero order, $n=0$, has the largest intensity. The distance between maxima for a specific wavelength increases as λ increases. Thus, red wavelengths are more diffracted than blue.

Note that the maximum intensity is equal to I_0 since $(\sin u/u)^2 = 1$ for $u = 0$. Zero intensities occur when $u = n\pi$, or

$$b \sin \beta = n\lambda. \quad (1.23)$$

1.3.2 The diffraction grating

We now have the tools to understand the diffraction grating. A diffraction grating consists of N single slits aligned as shown in Fig. 1.11. The spacing between the slits is a , and the width of each slit is b . Each slit will act a source with an intensity given by equation (1.22). These N emitters will produce interference according to equation (1.16). The net intensity is then simply interference caused by N slits, modulated by the diffraction pattern of one single slit

$$I = I_0 \left[\frac{\sin(N\pi a \sin \beta / \lambda)}{\sin(\pi a \sin \beta / \lambda)} \right]^2 \times \left[\frac{\sin(\pi b \sin \beta / \lambda)}{\pi b \sin \beta / \lambda} \right]^2. \quad (1.24)$$

Fig. 1.12 plots equation (1.24) as a function of wavelength λ and angle β .

Note that for each wavelength we obtain maxima at different angles except for zero order. This is the same for all wavelengths. Or in other words, we obtain spectra on each side of zero order. Also note that the location of the intensity maxima for each order increases for increasing wavelength. This effect is the opposite of what happens with a prism, where there are no spectral orders and blue light is more refracted than red.

1.3.3 Reflective gratings

In the previous section we looked at gratings that transmit light. The plane reflective grating³ may be thought of as a polished surface that has parallel grooves which have been made by a sharp diamond. The grating consists of narrow parallel mirrors, where each mirror acts a source of interference in the same manner as described above. See panel (1) of Fig. 1.13.

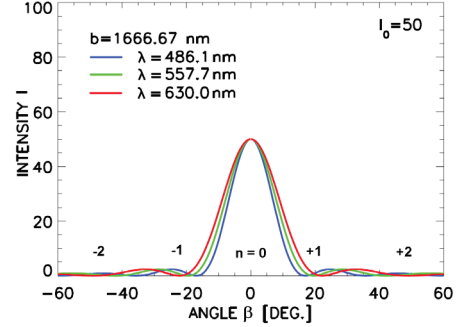


Figure 1.10. Diffraction as a function of angle β and wavelength λ . $b = 1666.67$ nm. $I_0 = 50$.

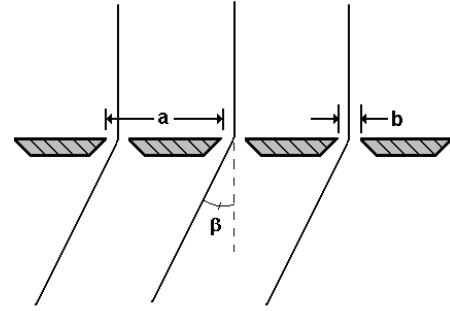


Figure 1.11. The diffraction grating.

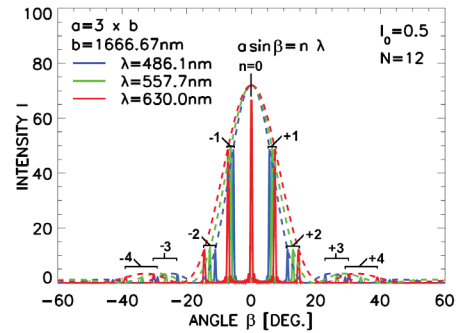


Figure 1.12. The diffraction grating pattern as a function of angle β and wavelength λ . $b = 1666.67$ nm and $a = 3b$. $I_0 = 0.5$ and $N = 12$.

The phase difference between each source due to path difference is sketched in panel (2) of Fig. 1.13

$$\begin{aligned}\sigma &= \frac{2\pi}{\lambda}(BC + AD) \\ &= \frac{2\pi}{\lambda}(a \sin \alpha + a \sin \beta)\end{aligned}\quad (1.25)$$

Maximum interference occurs at $\sigma = 2\pi m$, or
 $n\lambda = a(\sin \alpha + \sin \beta)$. (1.26)

Equation (1.26) is known as the *grating equation*, where n is the spectral order, α the incident angle, and β the diffracted angle of the grating. If $\alpha = \beta$, then the condition is called the *Littrow configuration*.

Note that so far most of the intensity is diffracted into the zero order. If the reflective surfaces are tilted at an angle ω_b to the grating normal \mathbf{k} , we have a *blazed* grating. ω_b is known as the blaze angle. Fig. 1.14 shows how a blazed grating may look like. Note that the phase difference between the rays S_1 and S_2 , as a result of the grooves, is the same as in equation (1.25). This means that the interference pattern is not shifted in angle, and that the grating equation still holds.

This is not the case for rays S_0 and S_1 . The blaze angle ω_b will introduce a shift of the diffraction pattern away from zero order. The phase difference is now written as

$$\sigma = \frac{2\pi}{\lambda}(BC + AD) = \frac{2\pi}{\lambda}(d \sin \varepsilon + d \sin \tau) = \frac{2\pi}{\lambda}a \cos \omega_b [\sin(\alpha - \omega_b) + \sin(\omega_b - \beta)]. \quad (1.27)$$

The diffraction pattern of the blazed grating can now be found by the same method as described above but using updated the phase differences for interference and diffraction, equations (1.25) and (1.27), respectively. The result is plotted in Fig. 1.15.

As seen in Fig. 1.15, a small blaze angle $\omega_b \approx 10^\circ$ will shift the diffraction pattern and the maximum intensity from zero to first order very effectively. Zero order is almost not visible.

Note from Fig. 1.14 that the grating becomes most efficient when $\alpha - \omega_b = \omega_b - \beta$, since then each surface will act as a small mirror, reflecting the rays with equal angles from the groove normal.

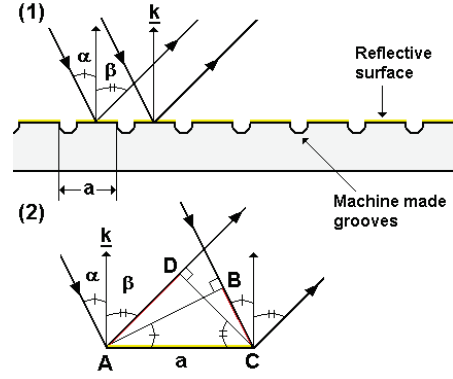


Figure 1.13. Panel (1): Ruled reflective grating. Panel (2): Rays with phase difference (red solid lines).

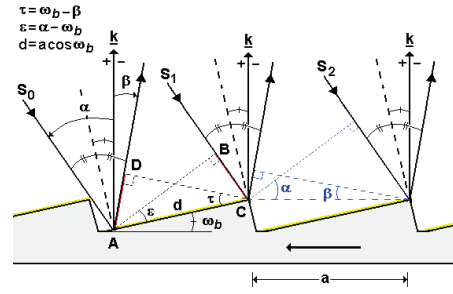


Figure 1.14. Blazed grating. Note that $\beta < 0$ due to sign convention.

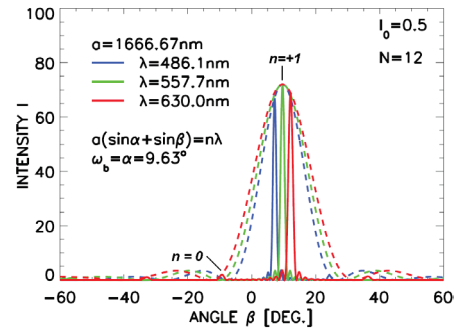


Figure 1.15. Blazed diffraction pattern.

An expression for the blaze wavelength λ_b can now be derived from the grating equation (1.26)

$$\begin{aligned}
 n\lambda &= a(\sin \alpha + \sin \beta) \\
 n\lambda &= 2a \sin \frac{1}{2}(\alpha + \beta) \times \cos \frac{1}{2}(\alpha - \beta) \\
 \alpha - \omega_b &= \omega_b - \beta \Rightarrow \alpha + \beta = 2\omega_b \\
 \alpha - \beta &= 2(\omega_b - \beta) = 2(\alpha - \omega_b) \\
 \lambda_b &= \frac{2a}{n} \sin \omega_b \cos(\alpha - \omega_b)
 \end{aligned} \tag{1.28}$$

A 600 grooves / mm grating with blaze angle $\omega_b = \alpha = 9.63^\circ$ gives a blaze wavelength $\lambda_b \approx 557.7 \text{ nm}$ for spectral order $n = 1$.

1.3.4 Overlapping spectral orders

One important behavior of a grating is that it produces multiple wavelengths at fixed diffraction angles as a function of spectral order. The overlapping wavelengths are a multiple of the factor $(1/n)$ from the grating equation (1.26). Fig. 16 shows the setup. A beam of white light hits the grating at an angle α to the grating normal \underline{k} . Three diffracted rays are shown as a function of wavelength and diffraction angle β .

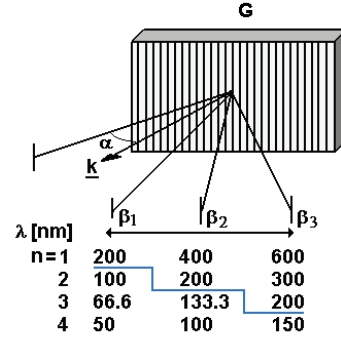


Figure 1.16. Spectral overlapping orders of a reflective grating (G). \underline{k} is grating normal and α incident angle. The three diffracted rays have angles β_1 , β_2 and β_3 .

Air at ground level absorbs light below $\sim 190 \text{ nm}$. Wavelengths below the blue line in Fig. 1.16 are therefore not detected.

If you, for example, need to make a first order measurement in the visible spectral range 400 to 600 nm, then you will need to block out second order overlapping wavelengths from 200 to 300 nm. This is in the ultraviolet part (UV) of the electromagnetic spectrum. A cut off glass filter will be needed in order to block out this range, if your detector is sensitive to UV.

1.3.5 Resolving power

The resolving power R is a theoretical concept defined as

$$R \equiv \frac{\lambda}{\Delta\lambda} \tag{1.29}$$

It is a measure of a grating's ability to separate adjacent spectral lines. Two spectral lines are considered resolved if the distance between them is such that the maximum of one falls on the first minimum of the other. This is known as the Rayleigh criterion.

The peak intensity of a grating occurs when the phase difference is $\sigma = 2\pi m$. The closest minimum occurs when $\Delta\sigma = 2\pi / N$. Differentiating equations (1.12) and (1.17) gives

$$d\sigma = \frac{2\pi}{\lambda} a \cos \beta d\beta \tag{1.30}$$

and

$$a \cos \beta d\beta = n d\lambda . \quad (1.31)$$

Therefore,

$$d\sigma = \frac{2\pi}{\lambda} n d\lambda = \frac{2\pi}{N} . \quad (1.32)$$

Rearranging equation (1.32) gives us

$$R \equiv \frac{\lambda}{\Delta\lambda} = n N . \quad (1.33)$$

For example, a resolving power of $R=(600 \text{ grooves/mm}) \times 12 \text{ mm} = 7200$, will result in $\Delta\lambda=0.07 \text{ nm}$ at $\lambda=500 \text{ nm}$ for $n=1$.

1.3.6 Angular dispersion

The angular dispersion is defined as $\frac{d\beta}{d\lambda}$. Differentiating the grating equation (1.26) we simply obtain

$$\frac{d\beta}{d\lambda} = \frac{n}{a \cos \beta} . \quad (1.34)$$

It is a measure on how rays spread out in a diffracted angle per unit wavelength, and it plays a key role in the calculation of the instrumental bandpass. Note that the angular dispersion increases with order.

1.3.7 Grating efficiency

The theoretical efficiency E_λ^n of a grating may now be found as the integral factor between the intensity at spectral order n and the total diffraction pattern.

Or since $I = I_\lambda(\beta)$ then

$$E_\lambda^n = \left[\frac{\int_{\beta_n - \Delta\beta}^{\beta_n + \Delta\beta} I_\lambda(\beta) d\beta}{\int_{-\pi/2}^{\pi/2} I_\lambda(\beta) d\beta} \right] \times 100 . \quad (1.35)$$

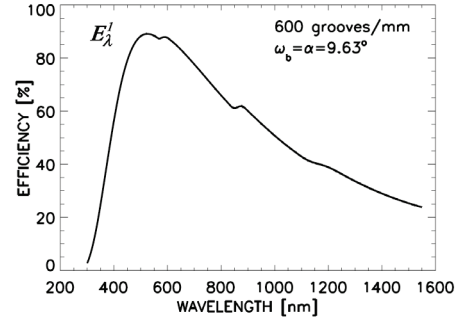


Figure 1.17. Theoretical efficiency for a grating with 600 grooves / mm and a blaze angle $\omega_b = \alpha = 9.63^\circ$. The blaze wavelength $\lambda_b = 557.7 \text{ nm}$ for $n = 1$.

Fig. 1.17 shows a typical efficiency curve E_λ^1 for a 600 grooves / mm grating, blazed at 9.63° . For each wavelength λ and spectral order n we obtain β_n from the grating equation. $\Delta\beta$ is the angle of peak interference down to first minimum from equations (1.30) and (1.32)

$$\Delta\beta = \frac{\lambda}{Na \cos \beta_n} . \quad (1.35)$$

The grating efficiency also depends on which coating material is used. For example, gold reflects light 10-15% better than aluminum in the near-infra red region of the spectrum.

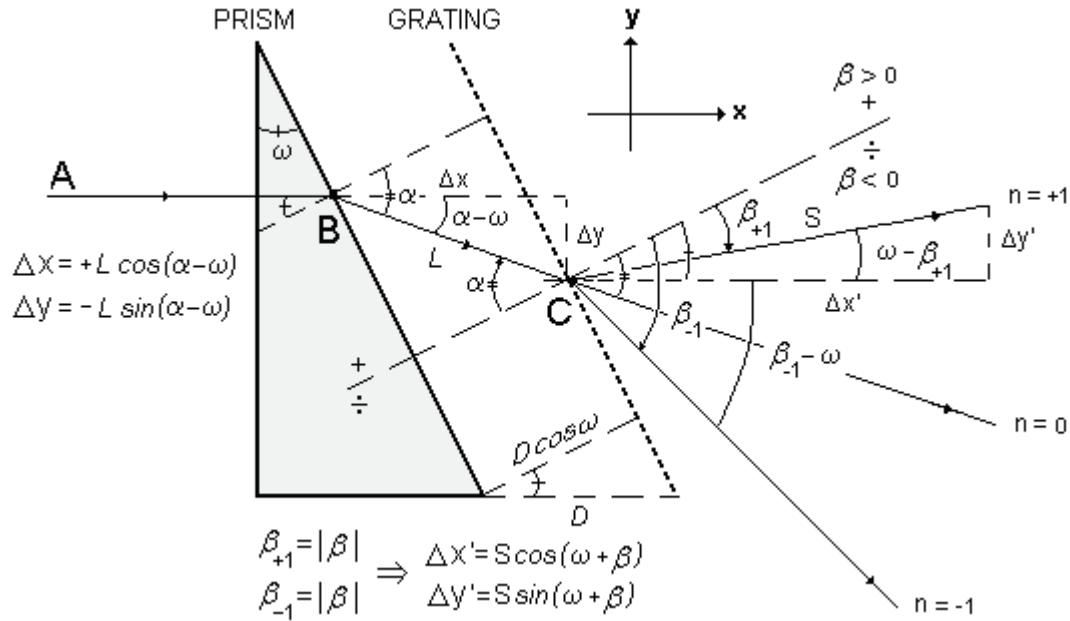


Figure 1.18. The GRISM configuration with grating sign convention and xy-coordinates.

1.4 The GRISM

A prism stacked in series with a transmission grating forms a diffractive element known as the GRISM or the Carpenters prism. Fig. 1.18 shows a wedge prism and a transmission grating separated a distance D in the x -direction (air gap). The ray from A to B is parallel with the x -axis. It does not refract since it hits the prism head on. At point B it ray will refract according to Snell's law

$$n_p \sin \omega = \sin \alpha. \quad (1.36)$$

The prism index of refraction n_p is given by Cauchy's formula as

$$n_p = A_1 + \frac{B_1}{\lambda^2}, \quad (1.37)$$

where A_1 and B_1 are constants depending on prism material. The distance from point B to C is now calculated as

$$\begin{aligned} \Delta x &= L \cos(\alpha - \omega) \\ \Delta y &= -L \sin(\alpha - \omega) \end{aligned} \quad (1.38)$$

where $L = D \cos \omega / \cos \alpha$. At point C the refracted angle α becomes the incident angle to the grating. The grating equation may now be modified by equation (1.36)

$$n\lambda = a(\sin \alpha + \sin \beta) = a(n_p \sin \omega + \sin \beta). \quad (1.39)$$

The position of diffracted rays of length S from point C become

$$\begin{aligned} \Delta x' &= S \cos(\omega + \beta) \\ \Delta y' &= S \sin(\omega + \beta) \end{aligned} \quad (1.40)$$

Equations (1.36) – (1.40) form the basics for ray tracking as function of wavelength.

Fig. 1.19 and Table 1.1 show the result of a computer generated ray tracking for the first spectral order in the visible part of the spectrum (400 – 700 nm). The prism has a diameter of 50 mm and a wedge angle $\omega = 30^\circ$. The material is Barium Crown (BaK4). The grating has 600 lines / mm.

Note that the yellow - green wavelength close to $\sim 500 \text{ nm}$ passes straight through the GRISM, parallel to the x -axis. This *on-axis* effect of the GRISM is favorable due to the simplicity to stack additional optics on both the front and back side of it. Image quality will be preserved due to negligible *off-axis* effects.

The total spread in the diffracted angles of the spectrum is also less than using a grating alone. The latter is due to the fact that a prism disperses blue light more than red, whereas the grating diffracts red light more than blue. The net effect is a compact spectrum centered at the straight through wavelength λ_s of the GRISM, or when $\alpha = -\beta = \omega$.

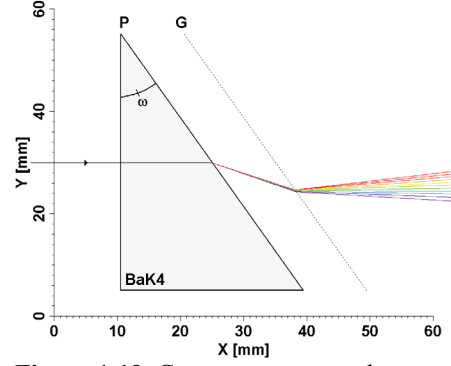


Figure 1.19. Computer generated ray tracking for GRISM. P is wedged prism with $\omega = 30^\circ$. G is 600 lines/mm grating. First order spectral range: 400 – 700 nm.

Wavelength λ [nm]	Refractive index n_g (BaK4)	Incident angle α [deg.]	Diffracted angle β [deg.]	GRISM angle $(\omega - \beta)$ [deg.]
400	1.602	+53.23	-34.13	-4.13
492	1.591	+52.70	30.02	0.02
700	1.580	+52.18	-27.71	+8.29

Table 1.1 GRISM data table for Fig. 19. Wedge angle $\omega = 30^\circ$. n_g is refractive index of BaK4. Grating has 600 lines/mm with incident angle α and diffractive angle β . The GRISM angle $(\omega - \beta)$ is relative to the x -axis.

The modified grating equation (1.39) becomes

$$n\lambda_s = a(n_p - 1)\sin \omega. \quad (1.41)$$

Furthermore, by differentiating equation (1.39) we obtain

$$n \frac{d\lambda}{d\beta} = a \sin \omega \frac{dn_p}{d\beta} + a \cos \beta = -2a \sin \omega \frac{B_1}{\lambda^3} \frac{d\lambda}{d\beta} + a \cos \beta \quad (1.42)$$

Rearranging gives the *angular dispersion* of a GRISM

$$\frac{d\beta}{d\lambda} = \frac{n + 2aB_1 \sin \omega / \lambda^3}{a \cos \beta}. \quad (1.43)$$

The term $2aB_1 \sin \omega / \lambda^3 > 0$, which implies that the GRISM has increased angular dispersion compared to using a grating alone.

2. SPECTRAL DESIGNS

2.1 The Ebert-Fastie spectrometers

In 1889 Herman Ebert⁵ described a spectrometer using a spherical mirror and a plane reflecting grating. Also known as a Plane Grating System (PGS). His sketch is reproduced in Fig. 2.2 and shows the instrument as a photographic spectrograph.

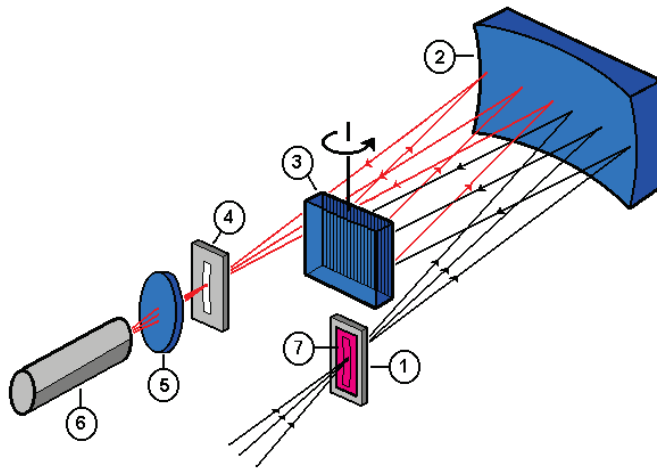


Figure 2.1. The Ebert-Fastie spectrometer: (1) curved entrance slit, (2) concave mirror, (3) plane reflecting grating, (4) curved exit slit, (5) collector lens, (6) photon counting detector, and (7) order sorting cut-off filter.

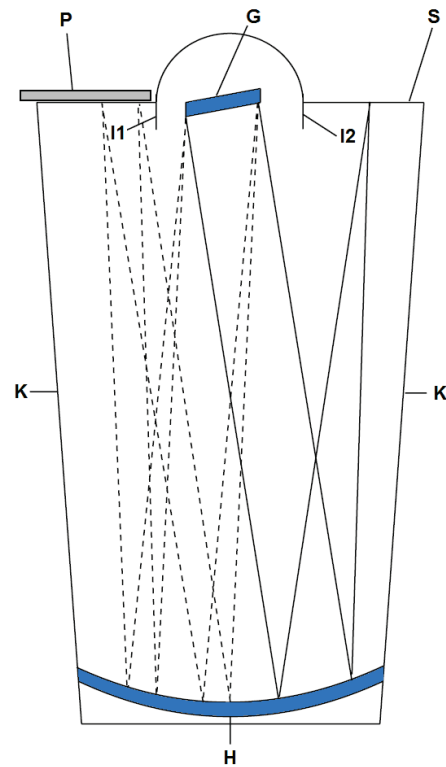


Figure 2.2. Reproduction of Ebert's sketch of his spectrometer. *I1* and *I2* are light baffles, *P* photographic plate, *G* grating, *S* entrance slit plane, *K* side plates, and *H* spherical concave mirror.

The light from the entrance slit *S* is first reflected by the right half part of the concave mirror *H*. The grating *G* is then illuminated by a parallel beam of light, since the entrance slit is located in the focus plane of the mirror. Secondly, the diffracted rays from the grating are focused by the left half part of the mirror on to the photographic plate *P*.

The grating may also be rotated to sweep an angular interval. The photographic plate is then replaced by an exit slit and a photon counting detector. As illustrated in Fig. 2.1, when the grating turns, the image of the entrance slit is observed at the exit slit in different wavelengths. This optical system is known as the *Ebert monochromator*.

In 1952 William G. Fastie improved the performance of the instrument, using curved slits instead of straight slits to reduce aberrations from the mirror.^{6,7} Problems of astigmatism became less, and the resolution of the instrument was improved. The instrument is therefore named the *Ebert-Fastie spectrometer*.

Note that the design only uses two active optical elements, which makes it ideal for low light applications such as airglow and faint auroras. Fig. 2.3 shows a 1m focal length auroral Ebert-Fastie. It is based on the original construction by Fastie at John Hopkins

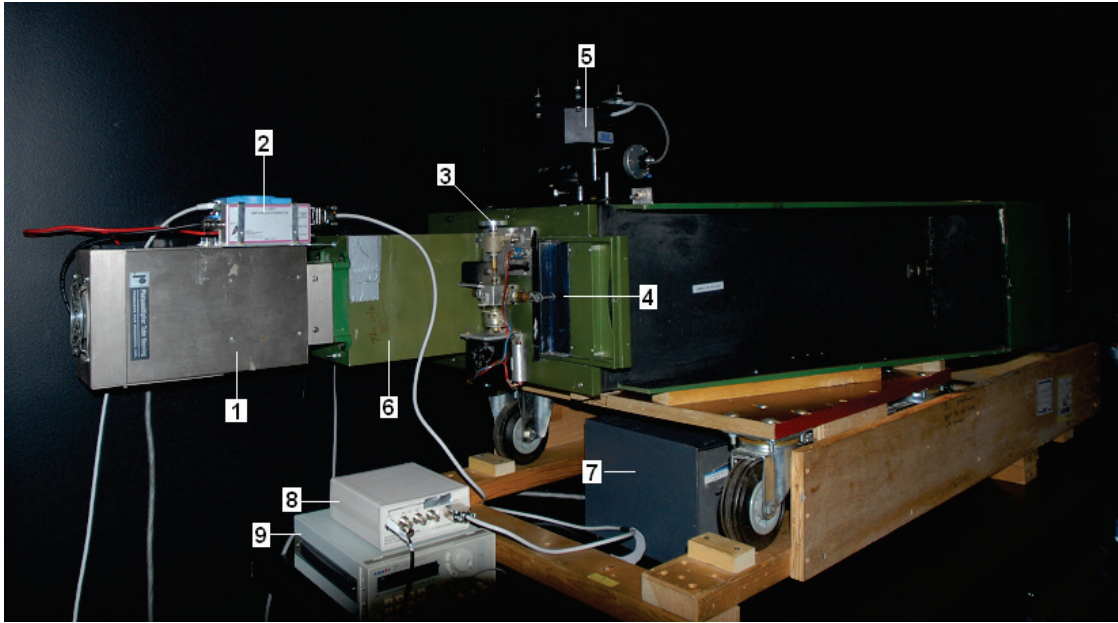


Figure 2.3 The 1m Green Ebert – Fastie spectrometer. (1) Peltier cooler for photomultiplier tube, (2) Pulse Amplifier and Discriminator (PAD) with High Voltage (HV) controller, (3) slit width adjuster wheel, (4) order sorting filter in front of entrance slit, (5) grating motor system, (6) light pipe, (7) power supply to cooler, (8) power to PAD, and (9) DC supply to the HV controller.

University, Maryland in the beginning of the 70's. In 1978, it was transferred by Gulamabas Sivjee and Charles Sterling Deehr to the Auroral Station in Adventdalen, Svalbard, Norway, from the Geophysical Institute (GI), University of Alaska, USA. Two more followed in 1980. The instruments are named 1m Green, ½m Black and 1m Silver Bullet according to focal length and chassis colour. Furthermore, Ove Harang's ½m White was moved in 2004 from the Skibotn Observatory in Norway up to Svalbard. In 2007, they were all installed at the new Kjell Henriksen Observatory (KHO) on Svalbard. Data from these instruments are widely published and recognized.

Even though there are only two active optical components used in the Ebert-Fastie spectrometer, the image quality is limited due to off-axis aberrations such as coma, astigmatism, curved focal field (defocusing) and spherical aberrations.

The effect of coma is seen as an asymmetric recorded spectral line profile. The base is enlarged on one of the sides of the spectral line as a function of wavelength. Defocusing causes the exit slit image to be blurred. Both will degrade the bandpass, the spatial resolution and the signal to noise ratio. They are due to off-axis geometry of the PGS system. Spherical aberration of the mirror will also blur the exit slit image due the effect that edge versus center rays fail to focus to the same point. It can only be corrected by the use of aspheric optics. Even worse is astigmatism that produces two foci, the tangential and the sagittal foci, when using a spherical mirror off-axis. The net effect is a slit height magnification at the exit slit. A point at the entrance slit tends to be imaged as a line perpendicular to the dispersion plane. The effect is minimized by aspheric optics and curved slits.

The basic equations for an Ebert-Fastie spectrometer are derived in Appendix A.

2.2 The Czerny-Turner configuration

In 1930 M. Czerny and A. F. Turner⁸ described a spectrometer using two concave mirrors instead of only one as in the Ebert-Fastie design. Fig. 2.4 shows the optical layout.

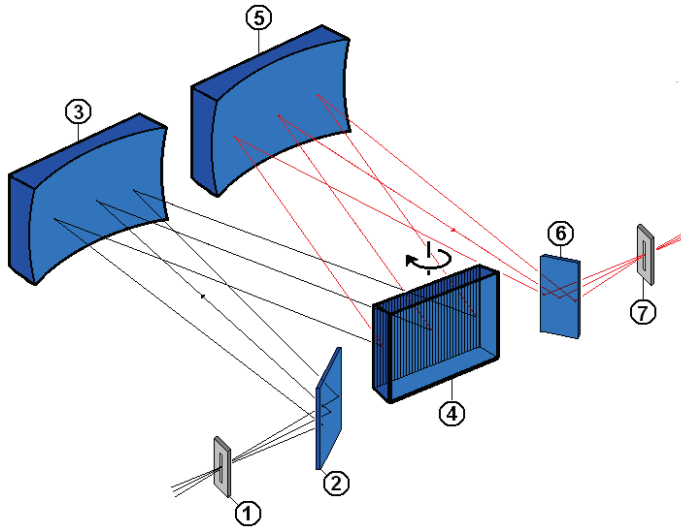


Figure 2.4. The Czerny-Turner configuration: (1) entrance slit, (2) flat mirror, (3) concave mirror (collector), (4) reflective grating, (5) focusing mirror, (6) flat mirror, and (7) exit slit.

It is clearly equal to the Ebert-Fastie configuration, but it is more flexible. The mirrors may now have different focal lengths, size and position compared to the more rigid design of the Ebert-Fastie.

The two flat mirrors labelled (2) and (6) in Fig. 2.4 are just used to make the design more compact. They are not used in the original layout⁸.

Off-axis aberrations still affect the image quality, but the use of two smaller mirrors instead of one large is cost efficient.

The price of high quality aspheric mirrors with minimum aberrations tends to increase dramatically with size or aperture. As a consequence, the Czerny-Turner configuration is one of the most used designs in spectroscopy.

2.3 The FICS concave grating spectrograph

Back in 1883 Henry Augustus Rowland invented the concave grating³ at the John Hopkins University. It both collimates and focuses the incident light beam. No extra mirrors or lenses are required. Today, holographic ion-beam etching techniques produce blazed concave gratings with high groove density and excellent image quality.

Fig. 2.5 shows a concave grating spectrograph made by Oriel® Instruments. The design is compact with no moving parts, and it has a high aperture ($f/\text{value} = 2$). The spectral range is 400-1100 nm. These qualities make it an ideal instrument for laboratory work, including lamp certification and absolute calibration of low light sources.

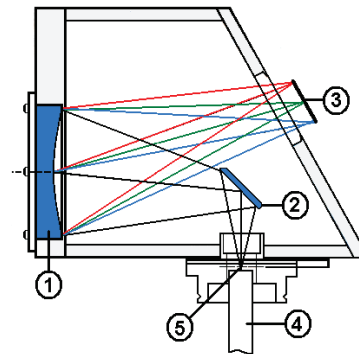


Figure 2.5. ORIEL Fixed Imaging Compact Spectrograph (FICS): (1) concave holographic grating, (2) flat mirror, (3) detector (CCD), (4) fiber bundle, and (5) entrance slit.

2.4 The GRISM spectrograph

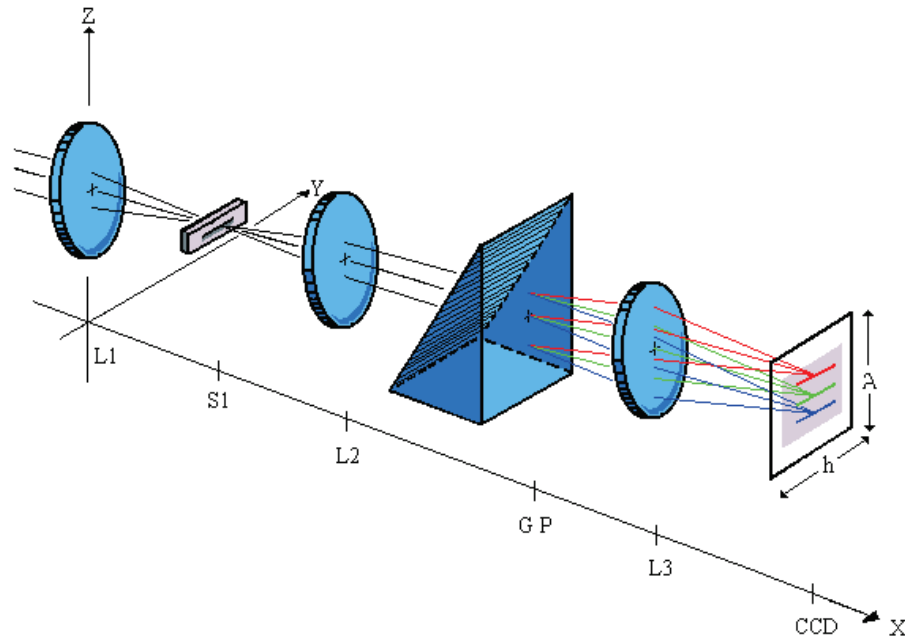


Figure 2.6. The GRISM spectrograph: *L1* front lens, *S1* entrance slit, *L2* collector lens, *G* grating, *P* prism, *L3* camera lens and *CCD* imaging detector.

One popular design is the use of a GRISM as dispersive element. As mentioned in section 1.4, this design makes it possible to obtain a straight through center wavelength parallel the optical axis of the system. Fig. 2.6 shows a typical configuration of a GRISM spectrograph. The optical elements may be stacked head on (on-axis). The front lens *L1* focuses light onto the entrance slit *S1*. Lens *L2* collimates the GRISM and lens *L3* focuses the diffracted light onto the image detector (*CCD*). The on-axis design reduces geometrical aberrations such as astigmatism and coma. The image quality is high compared to off-axis systems.

The grating may even be replicated or cemented straight onto the prism. No air cap is necessary as sketched in Fig. 1.18. A second prism may also be used to optimize the optical performance and protect the grating. A prism-grating-prism (PGP) element is shown in Fig. 2.7. An extra prism makes it possible to fine tune the straight through wavelength of the system⁹.

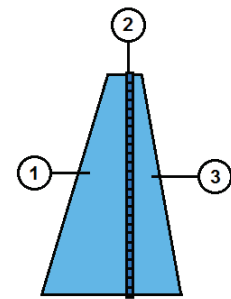


Figure 2.7. PGP-element. (1) Prism, (2) Grating and (3) Prism.

The company Specim Ltd. from Oulu in Finland has since the 90's produced a large number of PGP element based spectrographs for both industry and science. The only disadvantage of the PGP systems is the relative high number of optical elements needed compared to an off-axis mirror system such as Ebert-Fastie design. High aperture and minimum transmission losses through a minimum of optical elements are essential for low light applications.

2.5 The Auroral Spectrophotometer number 3 (SP3)

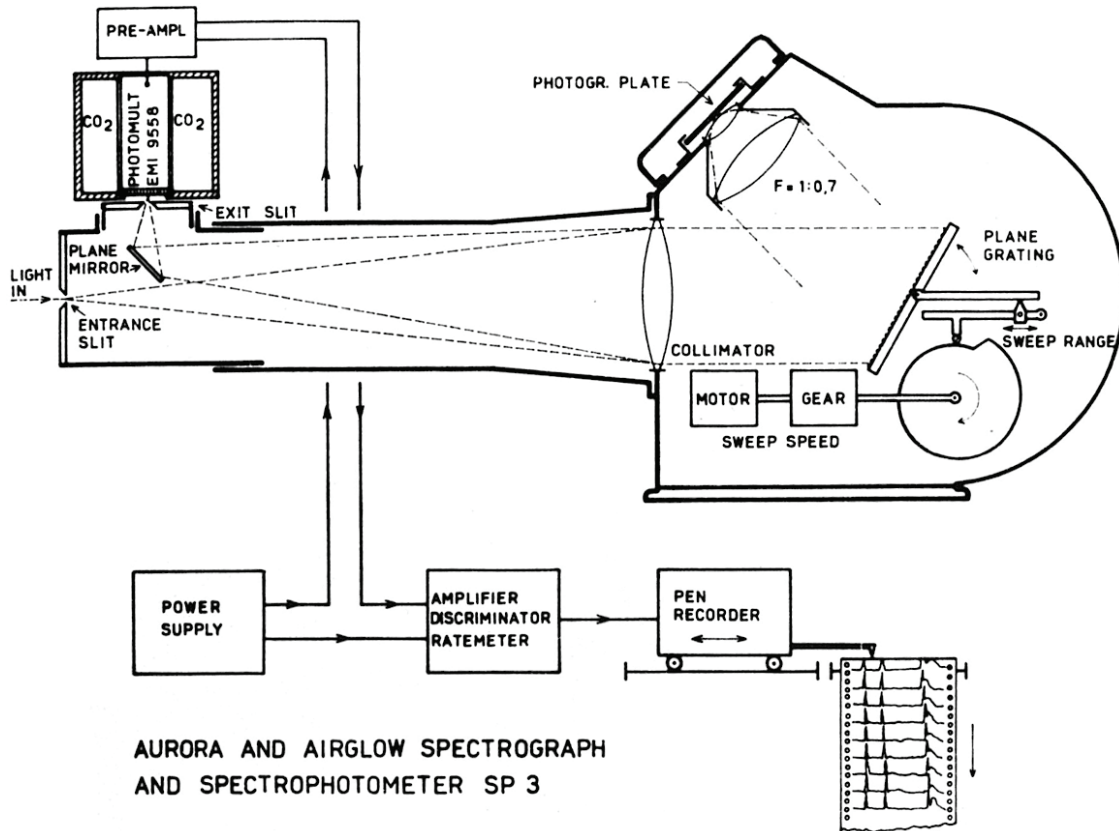


Figure 2.8. The SP3 combined spectrograph and Littrow spectrometer ¹⁰.

The last design presented in this chapter is the auroral spectrophotometer number 3 (SP3). The history of SP3 tells a story of auroral spectroscopy that started with Lars Vegard's spectrographs and the discovery of the proton aurora back in 1939. Vegard was former student of Kristian Birkeland and the founder of the Auroral Observatory in Tromsø. Later on the same institution, Anders Omholt and Leif Harang requested higher spectral resolution and sensitivity than Vegard's spectrographs, which used large quartz prisms instead of gratings as the key dispersive elements. The result was SP3 constructed in 1960 by Willy Stoffregen at the Ionospheric Observatory in Uppsala, Sweden. It was used extensively by his student Kjell Henriksen to study aurora, airglow and star spectra. In 1977 it measured for the first time Helium emissions in the polar cleft from Ny-Ålesund, Svalbard. It has even been used to obtain spectra of Barium released by rockets.

Fig. 2.8 shows the principal ray diagram and recording equipment of SP3 as originally designed by Stoffregen ¹⁰. The design has two modes. In the first mode, the collimator lens illuminates light from the entrance slit at an incident angle equal to the diffracted angle by the reflective grating. This is as mentioned in section 1.3.3 known as the Littrow condition. The collimator also focuses the light onto the exit slit. In front of the exit slit is a flat mirror that can be slightly bended to account for astigmatism.

As the grating sweeps in angular intervals, raw data from the cooled photon counting detector is plotted on the pen recorder. The result is stacked spectra with time. In the second mode, the instrument is used as a spectrograph with a photographic camera that looks down on the grating.

The instrument was found in late spring of 1993 stocked away in the garage of the Auroral Observatory in Tromsø. It was then completely covered with old scientific papers of the aurora, which most probably saved it from rust and degradation. The instrument was restored and the optical components realigned by the author under the supervision of his advisor Kjell Henriksen. Today, it is part of Svalbard Museum exhibition of the aurora.

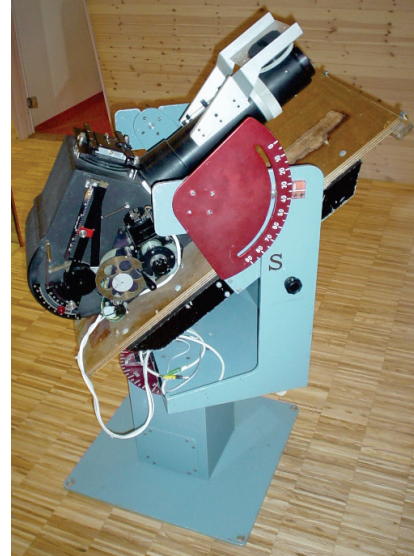


Figure 2.9. The SP3 at UNIS (2006).

Even though the SP3 is an old design from the 60's, it is still attractive for auroral studies with its dual modes, high spectral resolution and aperture. New control electronics and detectors have been developed since it was initially put to use. As with that the Ebert-Fastie spectrometers at KHO, the photon counting and computer electronics have been continuously upgraded since the late 80's. Or in other words, the optic designs stay more or less the same, while the field of optoelectronics has grown exponentially over the last decades.

3. SYSTEM OPTICS

In this chapter we will establish a general view on how to describe a spectrometer system. Parameters like the f /value, magnification and spectral bandpass will be explained.

3.1 Spectrometer optical diagram

An optical diagram is a standard way to trace rays through a spectrometer in an unrolled linear fashion. It visualizes a center cross section of the instrument perpendicular to the slits, parallel to the axis of refraction. Fig. 3.1 shows the layout of an optical diagram.

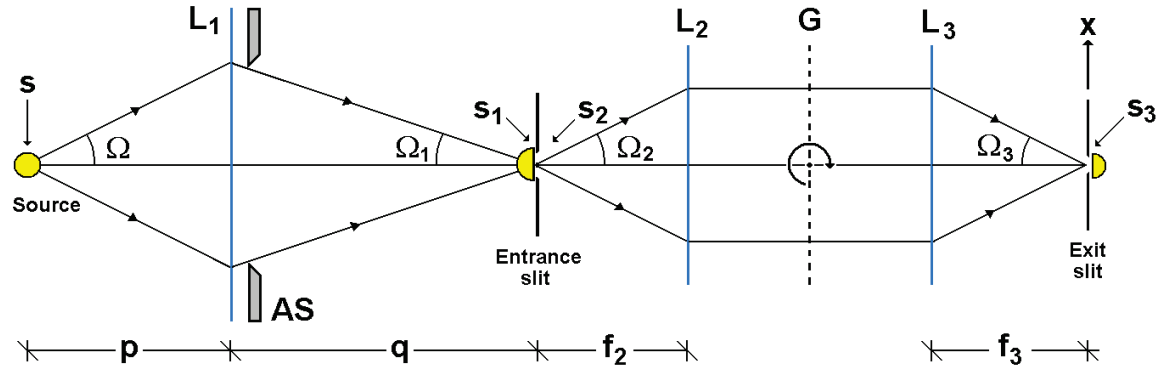


Figure 3.1. Reproduction in color of optical diagram of a spectrometer.⁴ AS is aperture stop, L_1 front lens, L_2 collimator lens or mirror, G grating, L_3 focusing lens or mirror, S area of light source, S_1 area of source image, S_2 area of entrance slit, S_3 area of the diffracted slit image (exit slit area), p object distance, q image distance of L_1 , f_2 focal length of L_2 , f_3 focal length of L_3 , and Ω half angles. The x -axis is parallel to the axis of diffraction, perpendicular to the slits.

Light from the source area S illuminates the front lens L_1 with half angle Ω and object distance p . S_1 is the imaged area of the source S that is focused onto the entrance slit plane by lens L_1 . Ω_1 is the half angle and q is the image distance of L_1 . Next, L_2 collimates the grating G with light that passes through the entrance slit area S_2 with half angle Ω_2 . L_3 focuses the diffracted light from the grating onto the exit slit plane with half angle Ω_3 . S_3 is then the area of diffracted entrance slit image. f_2 and f_3 are defined as the entrance arm length of L_2 and exit arm length of L_3 , respectively. In practice, these arm lengths corresponds to the focal lengths of L_2 and L_3 .

Note that the optical elements L_1 , L_2 and L_3 can be either lenses or mirrors as long as they are able to focus or collimate the beam. The grating may also be either transmitting or reflecting. The main points are that the entrance slit must be located in the focus planes of L_1 and L_2 , and that the rays form a parallel beam that illuminates the grating G . L_3 will then focus each parallel diffracted beam as a function of wavelength onto the exit slit plane.

The optical diagram is in other words, a general way to describe ray tracing through any spectrometer. It is a useful tool to use to calculate instrumental performance and evaluate different spectral designs.

3.2 Front optics

The front optics is in the optical diagram of Fig. 3.1, identified as L_I . The purpose is to image light from an object onto the entrance slit plane. L_I is not restricted to be only a single lens element. Any objectives such telescopes or even microscopes can be used as front optics. Nevertheless, the thin lens formula holds for most cases:

$$\frac{1}{f_1} = \frac{1}{p} + \frac{1}{q} . \quad (3.1)$$

f_1 is the focal length, p the object distance and q the image distance of L_I . Fig. 3.2 shows the 3 principal rays through a thin lens. Magnification of the lens is defines as

$$M_1 = \frac{q}{p} . \quad (3.2)$$

The speed of the lens is quantified by the ratio between focal length and the effective aperture, defined as the f/value or f/#

$$f/\#_1 \equiv \frac{f_1}{D_1} . \quad (3.3)$$

Likewise, numerical aperture of L_I is defined as

$$NA_1 \equiv \mu \sin \Omega_1 \quad (3.4)$$

Note that $\mu = 1$ for air.

From Fig. 3.3 the half angle of L_I for objects at infinity is

$$\Omega_1 = \arctan\left(\frac{D_1}{2f_1}\right) . \quad (3.5)$$

Applying equation (3.4) yields

$$NA_1 = \mu \sin \Omega_1 = \mu \sin\left(\arctan\left(\frac{D_1}{2f_1}\right)\right) \approx \mu \frac{D_1}{2f_1} , \quad (3.6)$$

which gives

$$f_1/\# \equiv \frac{\mu}{2NA_1} \quad (3.7)$$

Equation (3.7) shows us that $f/\#$ and NA are both a measure of light gathering power, and they are inversely proportional, i.e. a high NA results in a low $f/\#$.

3.3 f/value of a spectrometer

The f/value of a spectrometer depends on whether you observe the grating from the exit slit or the entrance slit. It also depends on the projected widths of the grating as seen from the entrance and exit slit. Namely,

$$\begin{aligned} W_2 &= W_g \cos \alpha \\ W_3 &= W_g \cos \beta \end{aligned} \quad (3.8)$$

where W_g is the width of the grating. See illustration in Fig. 3.4.

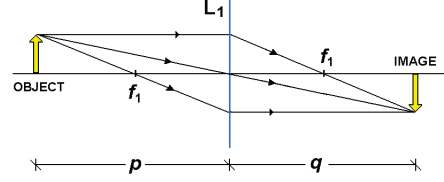


Figure 3.2. Single lens diagram of L_I with 3 principal rays. p is object distance, f_1 focal length and q is image distance.

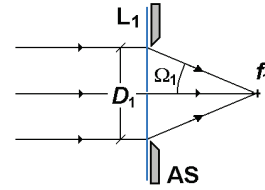


Figure 3.3. Single lens collimated. D_1 is diameter of lens, f_1 focal length and Ω_1 half angle.

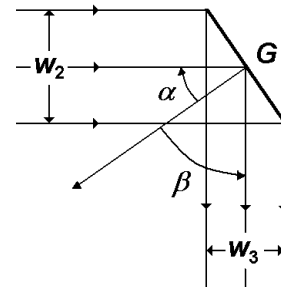


Figure 3.4. Illuminated grating.

Let us define D_2 and D_3 as the equivalent diameters as seen from the entrance and exit slit, respectively. If the height of the grating is H_g , then the following equations should hold assuming that the projected areas can be described as circular disks

$$\begin{aligned} \pi \left(\frac{D_2}{2} \right)^2 &= W_2 H_g = W_g H_g \cos \alpha & D_2 &= 2 \sqrt{\frac{W_g H_g \cos \alpha}{\pi}} \\ \pi \left(\frac{D_3}{2} \right)^2 &= W_3 H_g = W_g H_g \cos \beta & D_3 &= 2 \sqrt{\frac{W_g H_g \cos \beta}{\pi}} \end{aligned} \quad (3.9)$$

The f /values, as seen from entrance and exit, should then be

$$\begin{aligned} f/\#_2 &= \frac{f_2}{D_2} = \frac{f_2}{2 \sqrt{\frac{W_g H_g \cos \alpha}{\pi}}} \\ f/\#_3 &= \frac{f_3}{D_3} = \frac{f_3}{2 \sqrt{\frac{W_g H_g \cos \beta}{\pi}}} \end{aligned} \quad (3.10)$$

From equation (3.10) it is clear the f /value in and out of a spectrometer is not necessarily equal since it depends α and β , which implies that it depends on wavelength.

3.4 Magnification of the slits

The grating acts as a mirror along the entrance slit of height h . *Slit height magnification* is then simply the result of a lens with object distance f_2 and image distance f_3 . See Fig. 3.5. It then follows that

$$\frac{h}{f_2} = \frac{h'}{f_3} \Rightarrow h' = \left(\frac{f_3}{f_2} \right) \times h. \quad (3.11)$$

Slit width magnification is on the other hand

$$w' = w \times \left(\frac{\cos \alpha}{\cos \beta} \right) \times \left(\frac{f_3}{f_2} \right). \quad (3.12)$$

In the above equations the area of the entrance slit is $S_2 = w \times h$. The resulting de-magnified or magnified exit slit area is $S_3 = w' \times h'$. Equation (3.12) will be derived in full in Chapter 4.

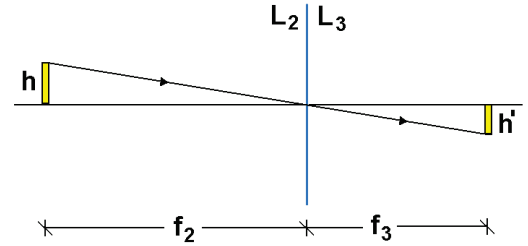


Figure 3.5. Lens magnification of the entrance slit height h of a spectrometer with collimator lens L_2 and detector lens L_3 .

3.5 Bandpass and resolution

Bandpass is the measure of an instrument ability to separate adjacent spectral lines. It is defined as the recorded Full Width at Half Maximum (FWHM) of a monochromatic spectral line. Fig. 3.6 shows the response of an instrument to a discrete monochromatic line. The half width of the line profile at intensity $\frac{1}{2} B$ defines our bandpass.

If F is the recorded spectrum, B the source spectrum and P the instrumental line profile, then F is given as the convolution between B and P

$$F = B \otimes P. \quad (3.13)$$

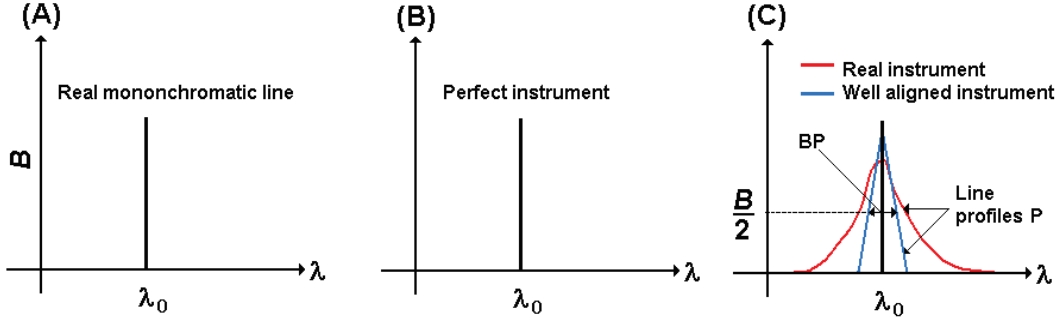


Figure 3.6. Effect of spectral resolution of monochromatic line at wavelength λ_0 and intensity B .⁴

The instrumental line profile is defined as

$$P(\lambda) \equiv P_1(\lambda) \otimes P_2(\lambda) \otimes \dots \otimes P_m(\lambda), \quad (3.14)$$

where each $P_i(\lambda) \forall i \in [1 \dots m]$ may be related to width of the slits, natural line width, resolution, alignment, diffraction effects, aberrations or quality optics etc.

The natural line width is neglected since it is usually not resolved. We assume no degradation due to aberrations and diffraction effects. The width of the instrumental profile is then determined by either the image of the entrance slit or the exit slit, whichever is greater. In other words, we assume $P(\lambda) = P_1(\lambda)$, and that the instrument is perfectly aligned.

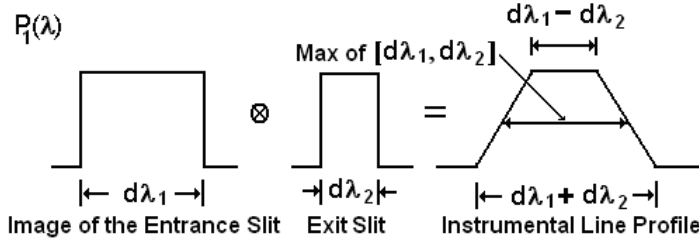


Figure 3.7. The instrumental profile due to the slits – a convolution of the entrance slits image with the exit slit.⁴

The shape of the P is controlled by the convolution of two rectangular functions with widths equal to the width of the entrance slit image, $d\lambda_1$, and the width of the exit slit $d\lambda_2$. The two slit functions are rectangular as illustrated in Fig.3.7. If the slits are perfectly matched then the line profile will be triangular and the FWHM equals half of the width at the base of the peak.

The spectral bandpass (BP) of the instrument is then calculated as the *linear dispersion* multiplied by the exit slit width

$$BP = FWHM \approx \frac{d\lambda}{dx} \times w'. \quad (3.15)$$

Applying equation (3.12), the inverse of equation (1.34) and $dx = f_3 d\beta$, then we finally obtain the bandpass as

$$\begin{aligned}
BP &= \frac{d\lambda}{d\beta} \times \frac{w'}{f_3} = \frac{a \cos \beta}{n f_3} w' = \frac{a \cos \beta}{n f_3} w \times \left(\frac{\cos \alpha}{\cos \beta} \right) \times \left(\frac{f_3}{f_2} \right) \\
BP &= \left(\frac{a \cos \alpha}{n f_2} \right) \times w
\end{aligned} \tag{3.16}$$

Note that bandpass depends mainly on L_2 and the entrance slit width. Bandpass varies as $\cos \alpha$, while dispersion varies as $\cos \beta$.

4. THROUGHPUT AND ETENDUE

The key question in this chapter is to find a way to quantify how much light is actually passing through a spectrometer? To answer this question, we need to define *etendue* and *throughput* of a spectrometer.

4.1 Definitions

The number of photons emitted from a source S per unit time into a solid angle Q is known as the flux Φ . Intensity I is defined as the flux per unit solid angle. Radiance B is the intensity through a unit surface area, or photons flux per unit area and solid angle.

According to the above, flux is given in units of

$$[\Phi] = \frac{\# \text{ photons}}{s}, \quad (4.1)$$

and intensity I has units

$$[I] = \frac{\# \text{ photons}}{sr s}. \quad (4.2)$$

Radiance B is then in units of

$$[B] = \frac{\# \text{ photons}}{cm^2 sr s}. \quad (4.3)$$

Fig.4.1 shows how the solid angle is defined.

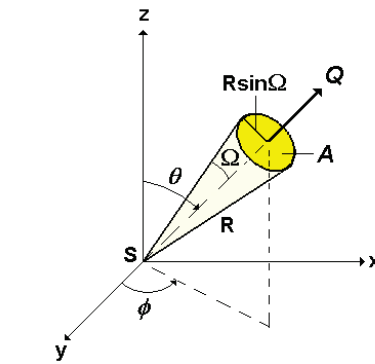


Figure 4.1. Solid angle.

4.2 Etendue

The field of view or the acceptance cone where photons are allowed to travel into will define how much light that is detected by an instrument. A useful quantity called the geometric extent (etendue) characterizes the ability of an optical system to accept light. It is a function of the area S of an emitting source and the solid angle Q its light propagates into or out of. It is defined as

$$G \equiv \iint dS dQ. \quad (4.4)$$

In our case we may write as $G = S \times Q$, where

$$Q = \frac{A}{p^2} = \frac{\pi(p \sin \Omega)^2}{p^2} = \pi \sin^2 \Omega. \quad (4.5)$$

The etendue is then simply

$$G = \pi S \sin^2 \Omega. \quad (4.6)$$

In terms of numerical aperture $NA = \mu \sin \Omega$

$$G = \pi S \left(\frac{NA}{\mu} \right)^2. \quad (4.7)$$

Equation (4.7) is useful when working with fibers. For example, a fused silica fiber with a diameter of $200 \mu m$ will have an etendue of

$$G = \pi S \left(\frac{NA}{\mu} \right)^2 = \pi (\pi \cdot 0.1^2) \cdot \left(\frac{0.22}{1} \right)^2 = 4.8 \times 10^{-3} mm^2 sr \quad (4.8)$$

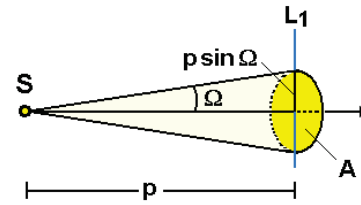


Figure 4.2. Front lens etendue.

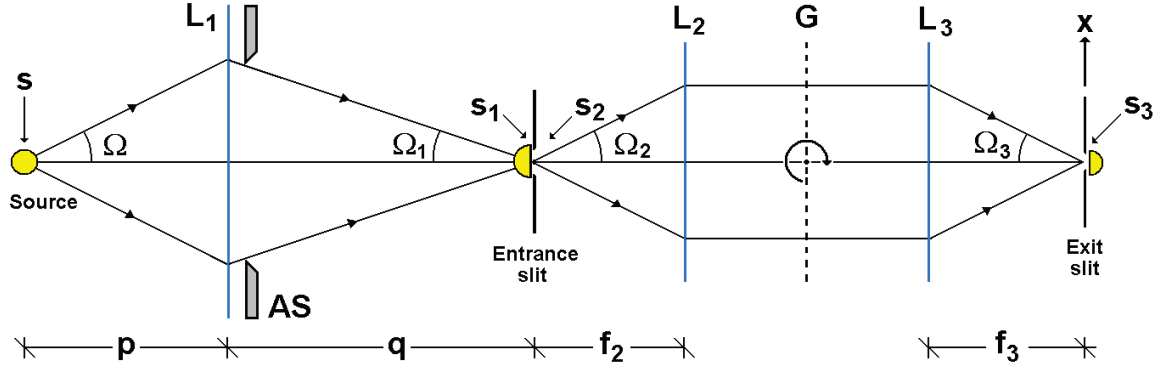


Figure 4.3. Optical diagram of a spectrometer.

Etendue may be viewed as the maximum geometric beam size an instrument can accept. It is a constant, and should be so throughout the instrument. Otherwise, we will lose light due to geometrical blocking etc. An instrument is said to be optimally constructed if the etendue is constant through each of the components used. Re-calling the optical diagram from Chapter 3, a spectrometer will be etendue optimized when

$$G = \pi S \sin^2 \Omega = \pi S_1 \sin^2 \Omega_1 = \pi S_2 \sin^2 \Omega_2 = \pi S_3 \sin^2 \Omega_3. \quad (4.9)$$

The optical diagram is reproduced in Fig. 4.3. Note that the front lens L_1 produces an image in the entrance slit plane. Only light from the image cross section that is defined by the width and height of the entrance slit will propagate into the instrument. It is important not to overfill the field of view of L_2 . The key etendue equation becomes

$$\pi S_2 \sin^2 \Omega_2 = \pi S_3 \sin^2 \Omega_3, \quad (4.10)$$

and it is defined as where the input etendue is equal to the output etendue of a spectrometer. Or in other words, the etendue is equal seen from both slits.

From the entrance slit the etendue is expressed as

$$G_2 = \left(\frac{G_A \cos \alpha}{f_2^2} \right) \times (w \cdot h), \quad (4.11)$$

where $G_A \cos \alpha$ is the illuminated area of the grating as seen from the entrance slit.

Correspondingly, at the exit slit

$$G_3 = \left(\frac{G_A \cos \beta}{f_3^2} \right) \times (w' \cdot h'). \quad (4.12)$$

Since $G_2 = G_3$, it follows that

$$\left(\frac{G_A \cos \alpha}{f_2^2} \right) \times (w \cdot h) = \left(\frac{G_A \cos \beta}{f_3^2} \right) \times (w' \cdot h'). \quad (4.13)$$

Using $h/h' = f_2/f_3$ then

$$w' = w \times \left(\frac{\cos \alpha}{\cos \beta} \right) \times \left(\frac{f_3}{f_2} \right). \quad (4.14)$$

This shows us that the etendue depends on the grating equation, size of slits, and the optics used. Thus, all parts of the spectrometer must be considered and complement each other.

4.3 Flux

Since etendue defines the maximum geometric cone and beam an instrument can accept, the photon flux through this cone must be a function of the radiance and the etendue

$$\Phi \equiv B \times G. \quad (4.15)$$

We see that

$$\begin{array}{ccc} \frac{\# \text{ photons}}{cm^2 s sr} \times cm^2 sr = \frac{\# \text{ photons}}{s} \\ \uparrow \quad \quad \uparrow \quad \quad \uparrow \\ [B] \times [G] = [\Phi] \end{array} \quad (4.16)$$

Equation (4.15) can now be used to evaluate photon fluxes in and out of a spectrometer system.

4.4 Throughput

Throughput is the usable flux at the exit slit, available to the detector. If we define B_2 to be the total radiance of the source entering the instrument at the entrance slit, then the total flux in is given as

$$\Phi_2 = B_2 \times G_2 = B_2 \times \left(\frac{G_A \cos \alpha}{f_2^2} \right) \times (w \cdot h). \quad (4.17)$$

At the exit slit the flux will be

$$\Phi_\lambda = B_\lambda E_\lambda^n T_\lambda \times G_3 = B_\lambda E_\lambda^n T_\lambda \times \left(\frac{G_A \cos \beta}{f_3^2} \right) \times (w' \cdot h'), \quad (4.18)$$

where E_λ^n is the efficiency of the grating at spectral order n at wavelength λ calculated in Chapter 1. B_λ is the spectral the radiance. The factor T_λ represents geometric losses (aberrations) and transmission factors of the optical components. $T_\lambda = 1$ for a perfect instrument

4.5 Stray light

4.5.1 Random stray light

The radiance of random stray light B_d due to scatter from different components such as mirrors, screws etc is directly proportional to the flux density Φ_2 / G_A or

$$B_d = C \times \left(\frac{B_2 (hw) \cos \alpha}{f_2^2} \right), \quad (4.19)$$

where C expresses the quality of the optics as function of random scatter. The random scatter flux at the exit is then

$$\Phi_d = B_d \times G_3 = \frac{CB_0 (hw) \cos \alpha G_A \cos \beta (h'w')}{f_2^2 f_3^2}. \quad (4.20)$$

The ratio of detected to random flux gives us an idea of the optical signal to noise ratio (S/N)

$$S / N = \frac{\Phi_{\lambda}}{\Phi_d} = \left(\frac{B_{\lambda}}{B_0} \right) \times \left(\frac{E_{\lambda}^n}{C} \right) \times \left(\frac{f_2^2}{h \cdot w} \right) \times \left(\frac{T_{\lambda}}{\cos \alpha} \right). \quad (4.21)$$

The main factors that control the optical signal to noise ratio in equation (4.21) is the quality of the optics as defined through the parameters E_{λ}^n , C and T_{λ} . Holographic gratings have a low C . This, one should make sure to use a grating with maximum efficiency in the wavelength region of interest, and use aberration corrected optics, if it is available.

4.5.2 Directional stray light

There are mainly 3 types of directional stray light sources:

1. Incorrect illumination of the spectrometer due to overfilled optics etc.
2. Re-entry spectra of unwanted orders / spectra that are focused onto the exit plane.
3. Grating ghosts and stray light due periodic machine ruling errors etc.

Solutions:

1. Use field lenses and aperture stops to illuminate correctly.
2. Use masks and side baffles on the grating. Tilt the detector.
3. Buy a new grating. As an ion-etched blazed holographic grating would be the recommended choice.

5. IMAGING SPECTROSCOPY

This chapter explains how we can image an object as a function of wavelength with high spectral and spatial resolution based on what we have learned in the previous chapters. As an example, we will use a GRISM spectrograph in hyperspectral image mode.

5.1 Short background

The rapid development of imaging detectors, especially the two-dimensional silicon Charge Coupled Device (CCD), has resulted in a new generation of imaging spectrographs. In contrast to the use of a photomultiplier tube as detector, the imaging detector provides information on the spatial distribution of the photon intensity along the entrance slit of the spectrograph. The ability of the CCD to image the entrance slit in different wavelengths and intensity has been used to reconstruct the image of the target object with high spectral and spatial resolution.

Note that during the last decade most of the effort in imaging spectroscopy (hyperspectral imaging) has been in the area of air- and space born sensors. A great deal of experience has been gained with these instruments and they have proven to be powerful tools in remote sensing.^{11, 12} As a result, the use of spectral imagers for industrial applications such as on-line spectral identification, sorting and inspection of objects has become increasingly important.^{13, 14, 15, 16, 17}

5.2 A low cost hyperspectral imager

A low cost portable instrument named DRONESPEX has been designed to demonstrate the technique of spectral imaging. A picture of the assembled system is shown in Fig. 5.1.

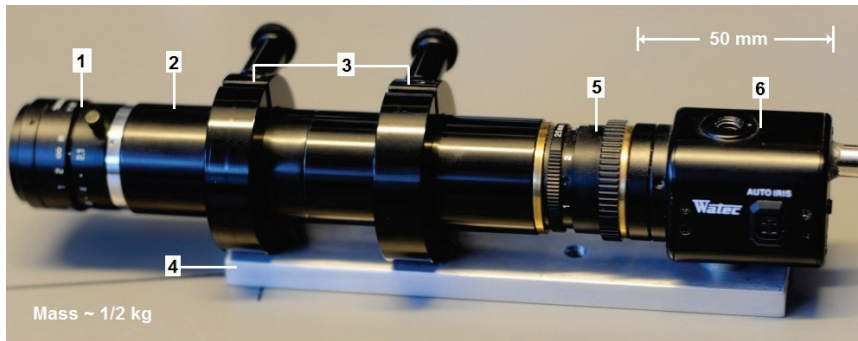


Figure 5.1. DRONESPEX.

A small portable low cost hyperspectral imager. (1) is front lens (objective), (2) aluminum barrel, (3) ring mounts, (4) mount rod, (5) camera lens and (6) CCD detector head.

The systems optical elements are shown in Fig. 2.6. The camera head (6) made by the company Watec LCL contains a 795×596 pixel CCD detector. The size of the CCD is $6.837 \times 4.9468 \text{ mm}^2$ with $8.6 \times 8.3 \mu\text{m}$ sized pixels. The spectral range is $400\text{--}700 \text{ nm}$ and the sensitivity is 0.00015 lux . The camera has both auto and manual and gain settings. The exposure time is 8.33 ms with a readout time of 31.67 ms (CCIR standard). The frames are stored on a digital video recorder. The dynamic range is 8 bits per pixel.

A handheld spectroscope made by Paton Hawksley Electronics Ltd. is fitted into the aluminum barrel (2). It contains a 600 grooves/mm transmitting holographic grating, a 30° degree Littrow dispersive prism, a $f_2 = 42 \text{ mm}$ focal length collimator lens, and a fixed sized slit. The lens is 12 mm in diameter. The slit is $h = 4.5 \text{ mm}$ high and $w = 25 \mu\text{m}$ wide. The grating is fixed parallel to the prism. The prism and grating normal coincide.

A standard C-mount $f_3 = 25 \text{ mm}$ focal length objective (5) is used to focus the diffracted and dispersed light from the grating/prism onto the CCD. The lens has manual iris control used to detect the background level of the CCD.

The field lens (1) or the front lens of the spectroscope is directly connected to the aluminum barrel (2). The barrel has standard inside C-threads. The spectroscope's entrance slit is located 17.5 mm into the barrel - positioned in the focal plane of the lens. The lens in Fig. 5.1 is a $f_l = 12 \text{ mm}$ focal length objective including variable focus and manual iris control. All mounts and adapters are found in the Mix and Match assemblies of the company Edmund Scientific (ES).

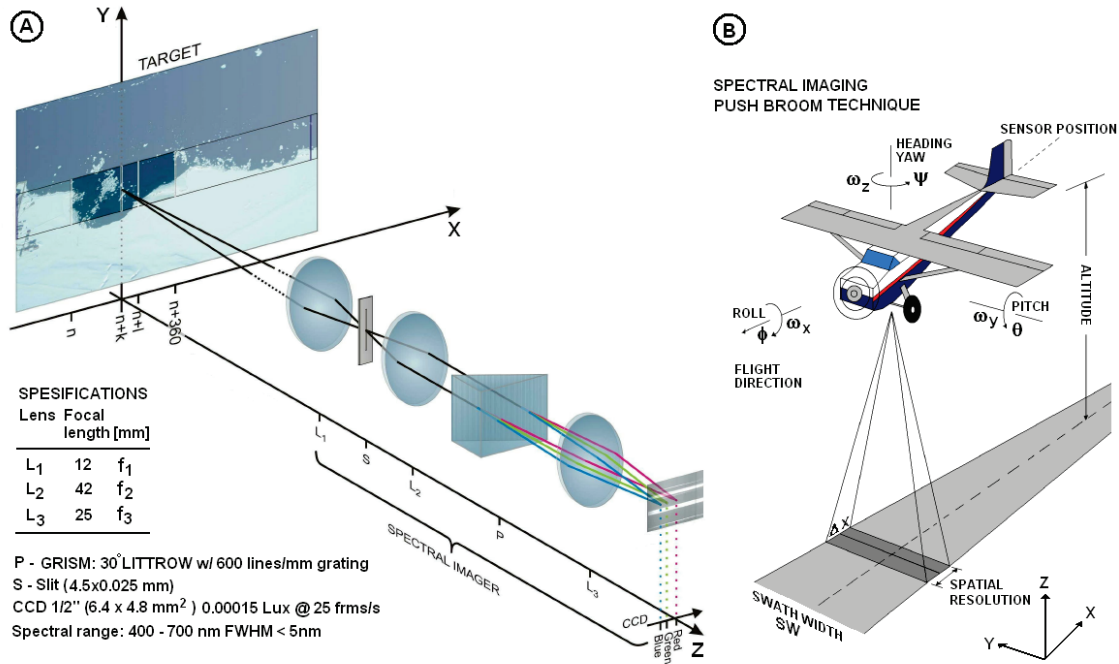


Figure 5.2. Basics of hyperspectral imaging. Panel (A): A 3 dimensional optical diagram illustrating the main principle. Each optical component is labeled with letters along the z – axis of the xyz – coordinate system. L_1 is front lens, S entrance slit, L_2 collector lens, P GRISM, L_3 camera lens and CCD imaging detector. The optical axis is parallel with the z - axis. The slit is located parallel to the y - axis. The detector is leveled in the xy - plane. Panel (B): Airborne carrier illustrating the push broom technique. The slit is oriented normal to the flight direction (x - axis). Precise estimation of the attitude of the plane requires a GPS for position, speed and heading, accelerometers for pitch θ , roll ϕ and yaw ψ , and gyros for turn rates (ω_x , ω_y , ω_z).

5.2 Main principle of hyperspectral imaging

Fig.5.2 panel (A) shows the optical components of the GRISM spectrograph as introduced in the previous section. The essential parts are again marked as S entrance slit, L_2 collector lens, P diffractive element – the grating / prism (GRISM), L_3 camera lens and CCD the detector. First, the main purpose of a spectrograph is that it generates images of the entrance slit as a function of color. The number of colored slit images on the CCD depends on the width of the entrance slit and the dispersive element's ability to spread – diffract the colors, and is directly connected to the spectral resolution of the instrument.

Secondly, focusing light from a target by a front lens (L_f) onto the slit - plane, forces the instrument to accept structure of the target along the slit. The resulting image (spectrogram) recorded by the *CCD* is the intensity distribution as a function of wavelength (color) and position along the slit. The spectrogram contains both spectral and spatial information along a thin track of the target object. In our case the target is snow cover land surface and ocean ice flakes.

Finally, in order to obtain the object's full spatial extent, it is necessary to sample the whole object. This requires that the instrument must be moved relative to the target. The whole idea is to record spectrograms for each track of the object as the image at the entrance plane is moved across the slit. Our movement is created by the airplane itself. See Fig.5.2 panel (B). We are in other word pushing / recording spectrograms as we fly over the ground target. A method to estimate the attitude of the airplane will be presented in Chapter 7.

Note that depending on the application, the relative movement between sensor and target may be obtained by rotating the whole instrument, use of scanning mirrors as front optics, or moving the target itself. The target may for example be located on a conveyor belt or on a sliding table.

5.3 Hyperspectral image formation and spatial resolution

As described above we need to push / record frames (spectrograms) as we fly over the target area in order to generate images as a function of wavelength.

In terms of mathematics, the intensity at each pixel of the spectrogram as we move over the target may be defined as a 2 dimensional array

$$u_k = u_k[\lambda, h], \quad (5.1)$$

where λ represents the wavelength axis on the detector, parallel to the x-axis of Fig.5.2. Correspondingly h is position parallel to the y-axis at the detector. The sample counter k separates the frames in time. The first column of the target image ($k=1$) is the integration of the spectrogram over the desired wavelength region of interest from λ_1 to λ_2

$$J_{\lambda_1 \rightarrow \lambda_2}[h, k] = \sum_{\lambda_1}^{\lambda_2} u_k[\lambda, h] \Delta\lambda, \quad (5.2)$$

where $\Delta\lambda = \lambda_2 - \lambda_1$ is the bandpass of the image. The next column in the target image is $k+1$. Note that $\Delta\lambda$ does not necessarily have to be equal to the spectral bandpass (*BP*) of the instrument. The relation $\Delta\lambda \geq BP$ must hold. Bandpass and spatial resolution must be considered in more detail.

Recall the angular dispersion of a GRISM is given by equation (1.43)

$$\frac{d\beta}{d\lambda} = \frac{n + 2aB_1 \sin \omega / \lambda^3}{a \cos \beta}. \quad (5.3)$$

Since $dx = f_3 d\beta$ then linear dispersion is defined as

$$\frac{d\lambda}{dx} = \frac{d\lambda}{f_3 d\beta} = \left(\frac{1}{f_3} \right) \times \left(\frac{a \cos \beta}{n + 2aB_1 \sin \omega / \lambda^3} \right). \quad (5.4)$$

Linear dispersion gives the spread in wavelength per unit distance in along the x-axis of the detector. Table 5.1 shows how the linear dispersion changes with wavelength for our instrument.

Wavelength λ [nm]	Refractive index n_g	Diffracted angle β [deg.]	Linear dispersion $d\lambda/dx$ [nm/mm]
300	1.61829	38.9872	37.9174
400	1.58942	33.6908	48.0392
500	1.57606	29.2111	53.9186
600	1.56880	25.1126	57.7198
700	1.56442	21.2360	60.3967
800	1.56158	17.5051	62.3734
900	1.55963	13.8757	63.8542

Table 5.1. Linear dispersion using a GRISM with $\omega = \alpha = 30^\circ$, grating groove spacing $a = 1666.667$ nm (a 600 lines / mm), spectral order $n = 1$, and a detector lens with focal length of $f_3 = 25$ mm. Cauchy's index of refraction constants are $A_1 = 1.5523$ and $B_1 = 5939.39$ nm for Borate flint glass. See Appendix B for more calculations.

It should be noted that the linear dispersion is about 16 nm larger at 400 nm if the prism is removed, using only the grating. Correspondingly, at 800 nm the difference becomes only 4 nm. The grism improves the linear dispersion compared to using only a grating, especially in the blue part of the spectrum. The straight through wavelength of the system is about 480 nm where $\alpha = -\beta$.

The spectral bandpass (BP) of the instrument is calculated as the linear dispersion times the exit slit width. Using equation (5.4) and (4.14) we obtain

$$BP = \left(\frac{w}{f_2} \right) \times \left(\frac{a \cos \omega}{n + 2aB_1 \sin \omega / \lambda^3} \right). \quad (5.5)$$

Table 5.2 shows the slit width magnification and the bandpass of our instrument.

Wavelength λ [nm]	Slit width magnification	Bandpass BP [nm]
300	0.66	0.63
400	0.62	0.74
500	0.59	0.80
600	0.57	0.82
700	0.55	0.83
800	0.54	0.84
900	0.53	0.85

Table 5.2. Slit width magnification and theoretical spectral bandpass of a grism spectrograph. The prism is of Borate flint glass with $\omega = \alpha = 30^\circ$. The transmitting grating has 600 lines / mm. The spectral order is $n = 1$. The lens between the entrance slit and the GRISM has $f_2 = 42$ mm as focal length. The detector lens has a focal length of $f_3 = 25$ mm. The entrance slit width is $w = 0.025$ mm.

From the above it is safe to assume that we can use 1 nm as bandpass of the spectral range from 400 to 700 nm. The number of spectral channels available is then 300 not 795.

The maximum number of color planes or spectral channels that a hyperspectral imager can produce depends on the bandpass and the spectral range of the instrument, not the number of pixels along the wavelength axis of the detector. In addition, the bandpass also affects the spatial resolution through the field of view along the x-direction of the flight.

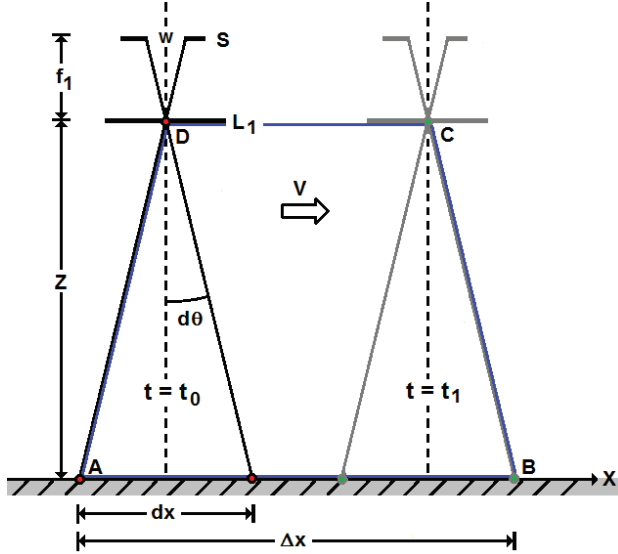


Figure 5.3. Field of view of slit as airplane moves with velocity v . S is slit with width w . L_1 is front lens with focal length f_1 . z is altitude above ground level.

The field of view of the slit as seen through the front lens along the direction of flight, defines the spatial resolution. It is illustrated in Fig. 5.3. The distance dx defines the ground segment seen by the instrument at time $t = t_0$

$$dx = \frac{z \times w}{f_1}, \quad (5.6)$$

where z is the distance to the target in meters, and f_1 is the focal length in millimeters of the front lens L_1 . Note that dx is connected to spectral bandpass BP through the width of the slit w .

During the exposure the airplane moves a distance $v \cdot (t_1 - t_0) = v \cdot \Delta t$.

The spatial resolution then becomes equal to the distance from point A to B

$$\Delta x = dx + v \cdot \Delta t. \quad (5.7)$$

The exposure time Δt of the images does not include the readout time of the detector, τ . The distance moved during readout $v \cdot \tau$ must be less than dx . If not the case, the instrument will miss samples of the target area (under sampling). This criterion may then be expressed using equation (5.6) as

$$\tau \leq \left(\frac{z \times w}{f_1 \times v} \right). \quad (5.8)$$

It is clear that spatial resolution depends on altitude, velocity of airplane, front optics and slit width. The latter also defines bandpass. These parameters are all connected and must be evaluated in order to achieve optimal performance.

Normal to the flight direction the resolution is calculated simply as

$$\Delta y = \frac{z \times h}{f_1 \times N}, \quad (5.9)$$

where h is as before the slit height in millimeters and N is the effective number of pixels along the slit image. Table 5.3 gives the spatial resolution of our instrument as a function of altitude using a sensor that pushes 25 frames per second, and an airplane carrier that flies at 200 km / hr.

$z[m]$	$dx[m]$	$\Delta x[m]$	$\Delta y[m]$	$SW[m]$
3000	6.25	6.71	8.65	1124.5
2500	5.21	5.67	7.21	937.3
2000	4.17	4.63	5.77	750.1
1500	3.13	3.59	4.33	562.9
1000	2.08	2.54	2.88	374.4
850	1.77	2.23	2.45	318.5
500	1.04	1.50	1.44	187.2

Table 5.3. Spatial resolution of for our GRISM spectrograph described above. The front lens has a focal length of $f_1=12\text{ mm}$. The speed of the plane is 200 km/hr (55.5 m/s). The entrance slit width is $w = 0.025\text{ mm}$. The sensor is a $\frac{1}{2}$ " CCD (320×240 pixels) at 25 frames per second. $N=130$ and $\Delta t = 8.33\text{ ms}$, $\tau = 31.667\text{ ms}$. SW is the swath width. Supplementary, see Appendix C.

The readout distance becomes $v \cdot \tau = 1.76\text{ m}$, which means that we have to fly on an altitude of at least 850 m according to Table 3. The altitude limit can be lowered by increasing the distance dx in equation (13) by selecting a shorter focal length on the front lens L_1 . The spatial resolution is $2.23 \times 2.45\text{ m}^2$ with a swath width of 318.5 m .

5.4. Example applications

A whole range of possible applications can be investigated using hyperspectral imaging. Each spectral image represents how effectively a target scene absorbs, reflects or scatters light at the selected centre wavelength. A set of spectral images will produce a unique spectral finger print for each object within a target scene. This type of data becomes ideal as input for sorting of objects by the use of image processing algorithms like classification.¹⁸

Data from our airborne campaigns in the arctic cover application such as mapping of

1. Ocean colour and algae's
2. Vegetation and rocks
3. Sea ice and leads
4. Snow cover.

In the following examples, a commercial twin engine aeroplane of type Dornier-228, operated by the company Lufttransport AS with base in Longyearbyen, Svalbard (78°N , 15°E), has been used as carrier of the instruments. The pointing direction of the imagers is 30 degrees to nadir (side view). For each mission a sliding door has been mounted on the plane, which can be opened during flight when approaching the target area. Position and speed is obtained by using the external GPS antenna of the plane. The attitude (pitch, roll and yaw) is estimated by using real-time data from 3 axis accelerometers and electronic turn rate sensors. A Kalman estimator is used to process the attitude data.

Fig. 5.4 shows a typical example of data from our latest airborne spectral imager over Longyearbyen (78°N , 15°E), Norway. The two Gray scaled image strips to the left are generated with a centre wavelength of 546.1 nm . The bandpass is 5 nm . The altitude is 1698 m . The ground speed is 184 km/hr . The spatial resolution is close to 2 m . Panel (A) shows the 546.1 nm image of the cleaning facility for coal owned by the local coal company. A successful unsupervised classification with 10 classes is presented in panel (B). The classification is based on only 30 colour planes (images).

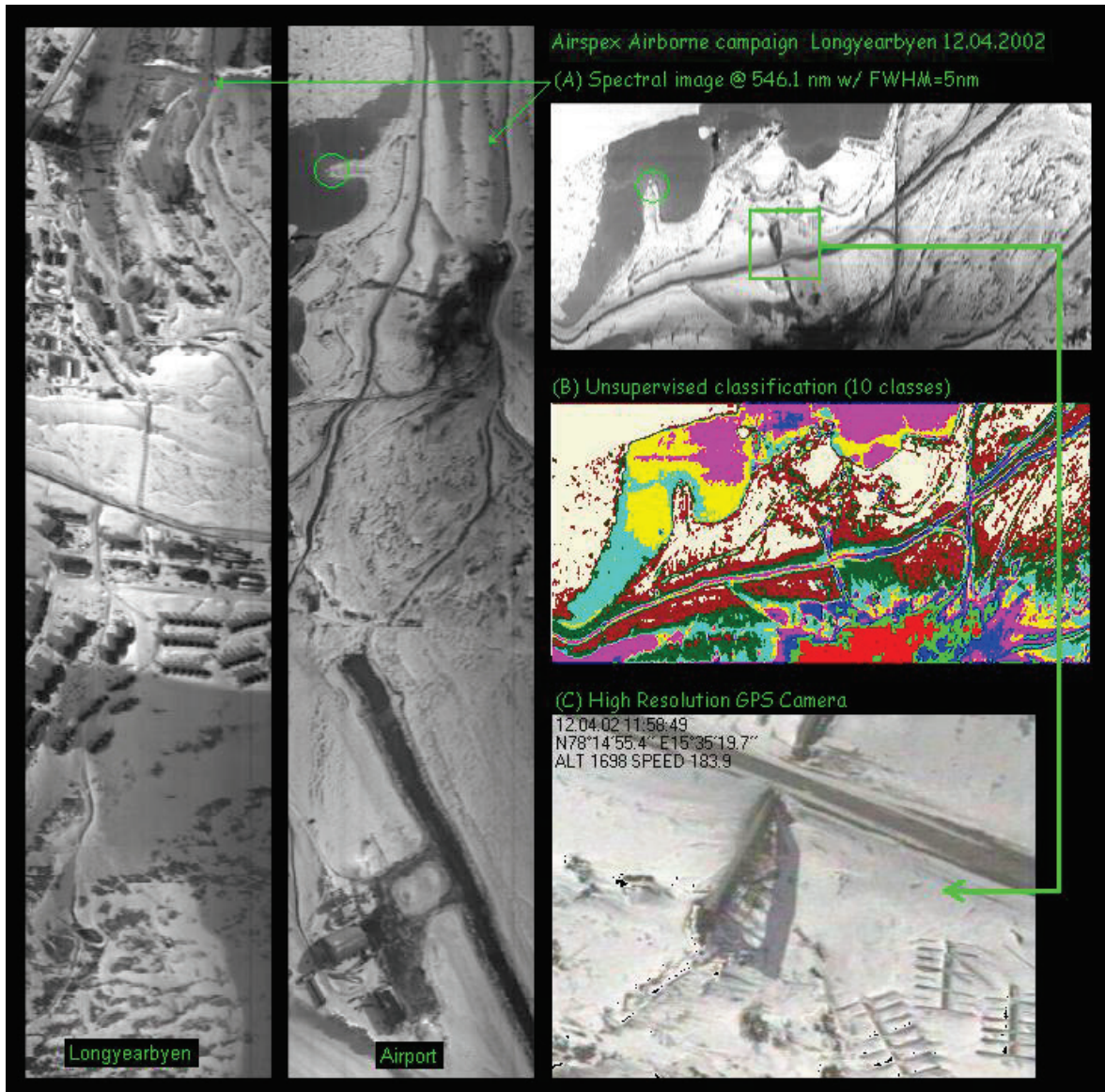
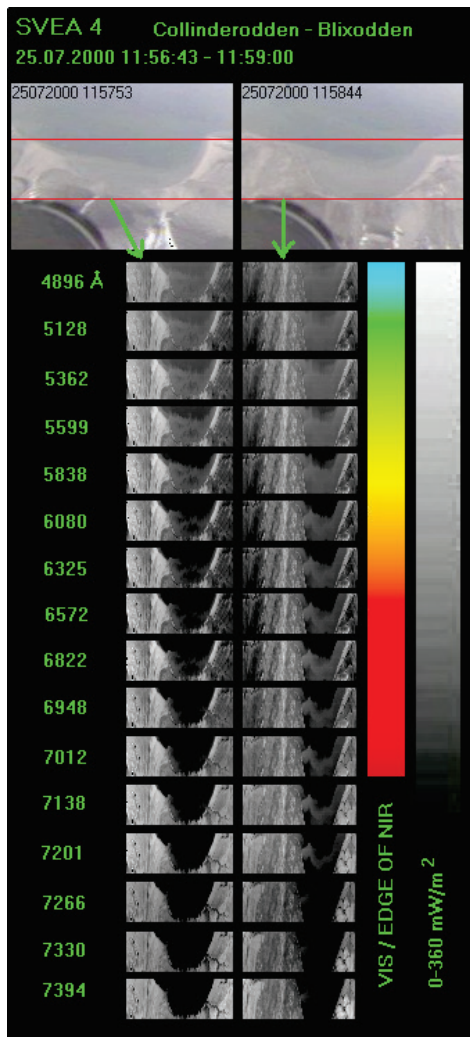


Figure 5.4. Hyperspectral imaging of snow cover and sea ice over Longyearbyen, Svalbard, 12.04.2002.

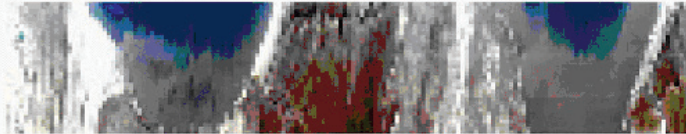
Note that the GPS position of the plane is found in the upper left corner of the high resolution camera image in panel (C).

The next example is shown in Fig. 5.5. The instrument provides images of ocean colour, which is related to the 3 basic components that effect the colour composition of sea water: Inorganic particles (e. g. mineral particles from fresh water run off – silt), phytoplankton with different pigment groups, and dissolved organic matter – yellow substance. The left panel shows hyperspectral data as a function of wavelength. The colour and grey scaled bars indicate the visible part of the electromagnetic spectrum and the intensity in absolute units, respectively. The altitude of observation is 3000 m. The spectral bandpass is 5 nm and the spatial resolution is approximately 7 m.



CLASSIFICATION

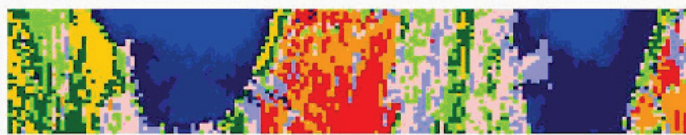
RGB true color composite: 633, 554, 490 nm FWHM=6 nm



RGB composite: 730, 578, 490 nm FWHM=6 nm



Unsupervised classification



Legend

Class_Names	Class_Names	Class_Names
Unclassified	Vegetation I	Rock I
Sea water	Vegetation II	Rock II
Silt I	Silt III	Rock III
Silt II	Freshwater silt	

Figure 5.5. Hyperspectral imaging of ocean colour, vegetation and rocks over the bays of Collinderodden and Blixodden close to the mining town of Svea, Svalbard, 25.07.2000

Focusing on the sea – mud interactions turns out to be very interesting. As we go up in wavelength the sea turn black and the mud (silt) source seems to appear. It may be that we can see the mud size distribution / morphology. This effect is not possible to see from the images taken by the digital camera (left panel's two top images).

Right panel of Fig. 5.5 shows the result of two RGB compositions and an unsupervised classification. The selected colour planes (the 3 images used in the compositions to represent Red, Green and Blue colours) bring out the areas dominated by vegetation and rocks. Furthermore, the classification reveals a more detailed map of 10 classes – separated in to sea water, 3 types of rock, 4 types of silt and 3 types of vegetation. Note that a supervised classification would improve and optimize the above results, but that requires in-situ observations on the ground. In many cases this can prove to be difficult due to logistics and access to the target area. Nevertheless, it is clearly evident that it is indeed necessary to combine in-situ observations to support and understand the airborne data better.

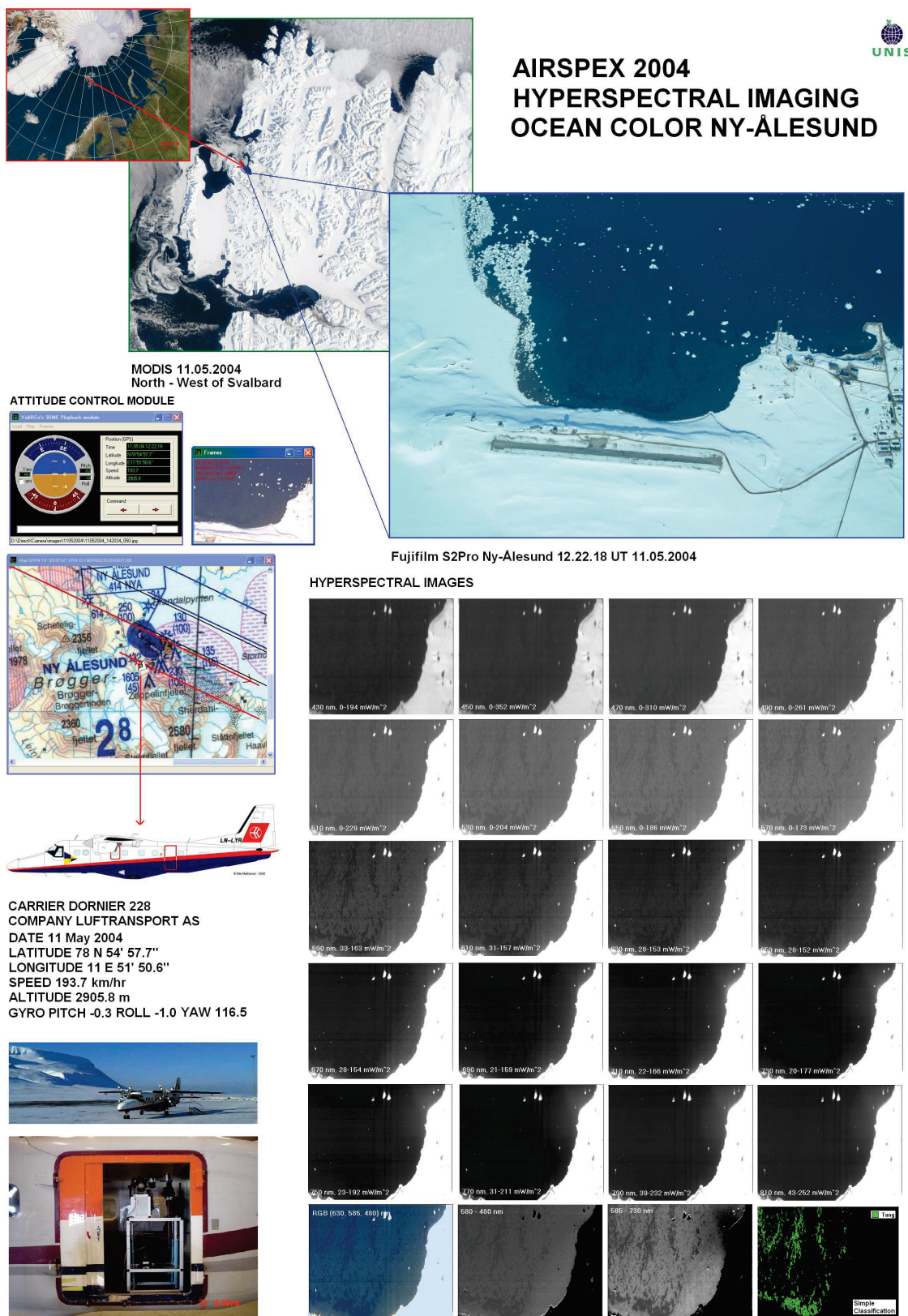


Figure 5.6. Hyperspectral poster from airborne campaign over Ny-Ålesund, Svalbard, 11.05.2004.

One more example of an airborne application is presented as a poster. See Fig. 5.6. A campaign was launched in May 2004 to study ocean color outside Ny-Ålesund (79°N, 12°E), Svalbard. Simultaneously with sampling of the airborne data, an image was taken by the Aqua - Terra satellite. The satellite image together with a high resolution color image (Fujifilm S2Pro digital camera with 28 mm objective) from the plane provided us with a wide coverage and a high spatial resolution of the target area, respectively. In addition, attitude and position tracks are shown to the left in Fig. 5.6.

The hyperspectral data are shown as a grey scaled image sequence throughout the visible part of the electromagnetic spectrum. The bandpass is 5 nm and the spatial resolution is close to 2 m. The composite RGB is made using the 480 nm (Blue), 585 nm (Green) and 630 nm (Red), respectively. Note from the difference between the images at 585 nm and 730 nm, structure on the bottom of the bay is clearly seen. Furthermore, a simple classification (Bayes) reveals a map of the brown algae population (mostly *Laminae* sp.).

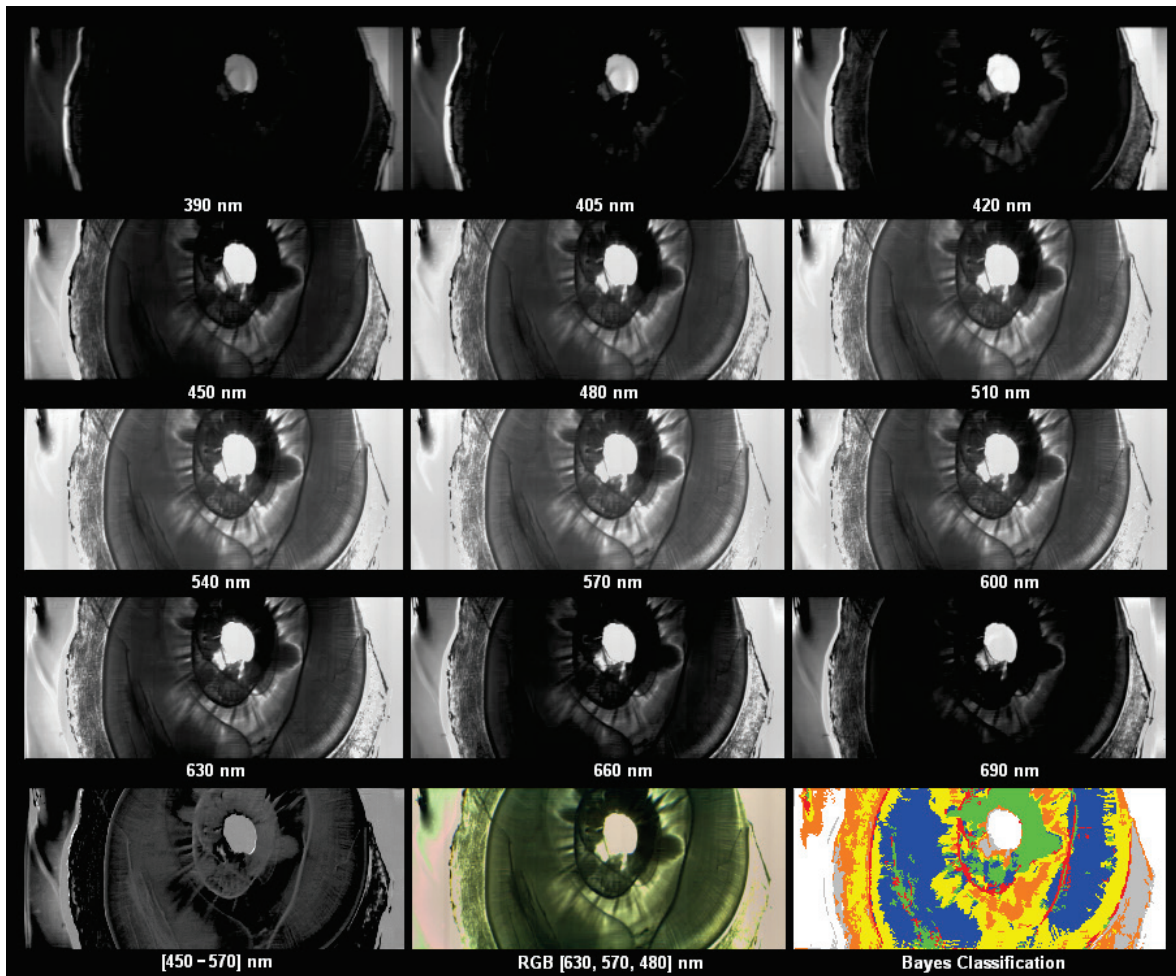


Figure 5.7 Hyperspectral data of sliced seal tooth. Front optics is microscope with 16 times magnification.

The final example application demonstrates that a spectral imager can also be used with a microscope as front optics. The relative movement between sample and instrument is provided by a sliding table. The sample is moved slowly under the microscope. Fig. 5.7 shows the spectral images from 390 to 690 nm of a sliced seal tooth. The bandpass of the images is 15 nm.

The slice is 1 *mm* thick. The magnification was set to 16 times on the microscope. The light source is located under the sample - light is transmitted through the sliced tooth. A processed image showing the difference between the 450 and 570 *nm* images reveals structure, and a composite RGB shows the colour image. Classification sorts out 8 different classes. Note that there is a small mix or miss classification. This is mainly due to cracks in the tooth and that the sample itself is glued to the sample glass (seen as a transparent area in the centre of the images). Nevertheless, the age rings of the tooth are clearly identified with the colour red. The seal was two years old.

5.5. What's next?

Optics is one of the oldest branches of physics. Spectroscopy started when Newton decided to celebrate the colors of life with a prism back in the late 1700. Even today optics is one of the fastest growing fields of physics, and is extending its use to other branches of science. The advances in electronics and new materials have resulted in detectors that are able to sense light outside the visible part of the spectrum. The recent development of Acoustic Optical Tunable Filters (AOTF) and Liquid Crystal Tunable Filters (LCTF) will revolutionize spectroscopy just as the arrival of blazed gratings did back in the 50's. UV imagers and near- to far infrared imagers will soon be available - even on the commercial market to a reasonable price. These spectral imagers will open our eyes to things we have never seen before.

6. CALIBRATION

The increasing number of low light level optical instruments operated in Svalbard (Longyearbyen, Barentsburg and Ny-Ålesund) for monitoring auroras and airglow phenomena emphasizes the need for establishing accurate calibration routines of international standard for both cameras and spectral instruments³².

6.1 Spectral narrow field of view calibration

This section reviews the mathematical frame work and the experimental setup of absolute calibration of narrow field of view spectral instruments. A presentation of the optical laboratory at UNIS is also given. The results of secondary standard lamp certification and brightness control are demonstrated.

6.1.1 Introduction

The throughput or the useable photon flux at the exit of an optical instrument depends first of all on input flux. Secondly, it depends on the instrument's ability to accept light (geometrical extent) and the quality of the optical components used (Chapter 4). The efficiency of each component, whether they are lenses, mirrors, gratings or filters, limits the sensitivity and the wavelength region of the instrument. The process of transforming the exit flux to electronic counts by the detector also involves loss of photons.

It is therefore necessary to calibrate the instrument against a source of known intensity in order to obtain the ratio of electronic counts out to the number of photons incident on the instrument. The source of the calibration could be a lamp or any other object with known spectral characteristics. In addition, a diffuse reflective surface is needed to make sure that the instrument's field of view is uniformly illuminated.

The following describes the theoretical concept and the experimental setup of sensitivity calibration using a Lambertian surface and a standard tungsten lamp. A method to transfer lamp certificates and a procedure to regulate source brightness of the screen without change in spectral shape is demonstrated.

6.1.2 The Lambertian surface

A surface that has a perfectly diffuse / matte property is Lambertian. The radiant intensity reflected in any direction is proportional to the cosine of the angle of the normal to the surface. This is known as Lamberts Cosine law¹⁹.

According to Fig. 6.1, the radiant intensity is expressed as

$$I_{\lambda\phi} = I_{\lambda 0} \times \cos\phi, \left[\frac{\# \text{ photons}}{s \text{ sr } \text{\AA}} \right] \quad (6.1)$$

where λ subscribes the wavelength and ϕ is the angle with respect to the normal. The total emission rate in the wavelength interval $d\lambda$ is

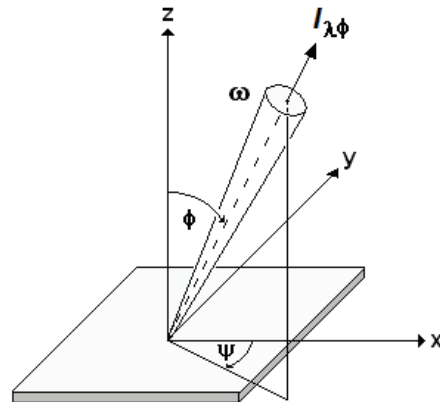


Figure 6.1. The radiant intensity of a Lambertian surface.

the radiant intensity integrated over the hemisphere

$$\Phi_{\lambda} = \int_0^{2\pi} \int_0^{\pi/2} I_{\lambda\phi} d\lambda \sin\phi d\phi d\psi. \quad (6.2)$$

Inserting Eq. (6.1) into (6.2) gives

$$\Phi_{\lambda} = 2\pi I_{\lambda 0} d\lambda \int_0^{\pi/2} \cos\phi \sin\phi d\phi. \quad (6.3)$$

$$= 2\pi I_{\lambda 0} d\lambda \int_0^{\pi/2} \frac{1}{2} \sin 2\phi d\phi. \quad (6.4)$$

Consequently,

$$\Phi_{\lambda} = \pi I_{\lambda 0} d\lambda. \left[\frac{\#photons}{s} \right] \quad (6.5)$$

The total emission rate for a Lambertian surface is independent of the angles ϕ and ψ .

Our surface (SRT-99-180) is made of Spectralon and is produced by the company Labsphere, Inc. The reflectance factors of the screen are nearly constant ($\rho_{\lambda} = 0.98$) throughout the visible and near infrared regions of the spectrum.

6.1.3 Experimental calibration setup

Fig. 6.2 shows the setup for the calibration. The tungsten lamp (FEL) is located a distance $z = 6.9 \text{ m}$ from the centre of the screen and the angle between the screen and lamp axis is given by α . $M_{o\lambda}$ is the known radiance (certificate) of the lamp in units of $[\#photons \text{ cm}^{-2} \text{ s}^{-1} \text{ \AA}^{-1}]$, initially obtained at a distance of $z_o = 0.5 \text{ m}$. The source for our calibration is then the screen, not the lamp itself.

The lamp is treated as a point source, radiating photons isotropically. The total number of photons that pass through a sphere with radius z_o must be the same for a sphere with radius z

$$4\pi z_o^2 M_{o\lambda} = 4\pi z^2 M_{z\lambda} \quad (6.6)$$

We assume no absorption of photons in the space between the spheres. The radiance that hits the screen is then simply

$$M_{z\lambda} = M_{o\lambda} \times \left(\frac{z_o}{z} \right)^2 \left[\frac{\#photons}{\text{cm}^2 \text{ s } \text{\AA}} \right]. \quad (6.7)$$

Correspondingly, the emission rate that enters the screen in terms of the radiant intensity $I_{s\lambda}$ is

$$I_{s\lambda} d\lambda \omega = I_{s\lambda} d\lambda \left(\frac{dA \times \cos\alpha}{z^2} \right), \quad (6.8)$$

where ω is the solid angle and dA is the illuminated area of the screen. The reemitted radiation of the screen is then

$$\Phi_{\lambda} = I_{s\lambda} d\lambda \rho_{\lambda} \left(\frac{dA \times \cos\alpha}{z^2} \right) = \pi I_{\lambda 0} d\lambda. \quad (6.9)$$

From Equation (6.1) we obtain
(6.10)

$$I_{\lambda\phi} = \rho_{\lambda} \left(\frac{I_{s\lambda}}{\pi} \right) \times \left(\frac{dA \times \cos \alpha}{z^2} \right) \times \cos \phi.$$

The effective illuminated area of the screen as seen by the instrument is $dA \times \cos \phi$, where ϕ is the angle between the optical axis and the screen normal. The radiance towards the instrument then becomes

$$B_{\lambda} = \frac{I_{\lambda\phi}}{dA \times \cos \phi} \left[\frac{\# \text{ photons}}{\text{s sr } \text{\AA cm}^2} \right] \quad (6.11)$$

$$= \rho_{\lambda} \left(\frac{I_{s\lambda}}{\pi} \right) \times \left(\frac{1}{z^2} \right) \times \cos \alpha. \quad (6.12)$$

From the inverse square law and Eq. (6.7) we know that

$$I_{s\lambda} = M_{z\lambda} z^2 = M_{o\lambda} z_o^2. \quad (6.13)$$

Note that the radiant exitance of the screen is by definition given as

$$M_{\lambda} = \pi B_{\lambda} \cdot \left[\frac{\# \text{ photons}}{\text{cm}^2 \text{s } \text{\AA}} \right] \quad (6.14)$$

Finally, we obtain from (6.12)

$$B_{\lambda} = \frac{M_{o\lambda}}{\pi} \rho_{\lambda} \left(\frac{z_o}{z} \right)^2 \times \cos \alpha. \quad (6.15)$$

Eq. (6.15) expresses the brightness of the screen as seen by the instrument in terms of the lamp and screen certificates, the distance between screen and lamp, and the angle of the screen to the lamp, α .

Note that as long as the field of view of the instrument is filled, neither the angle ϕ , nor the distance of the instrument to the screen matters. The changing size of the field of view at the screen compensates for the distance and the angle, ϕ .

6.1.4 The UNIS calibration lab

The described procedure requires a dark room to eliminate extraneous scattered light from the screen. Fig. 6.3 shows the new calibration laboratory at UNIS. The facility contains 3 rooms. The lamp room is separated from the screen room by a baffled door.

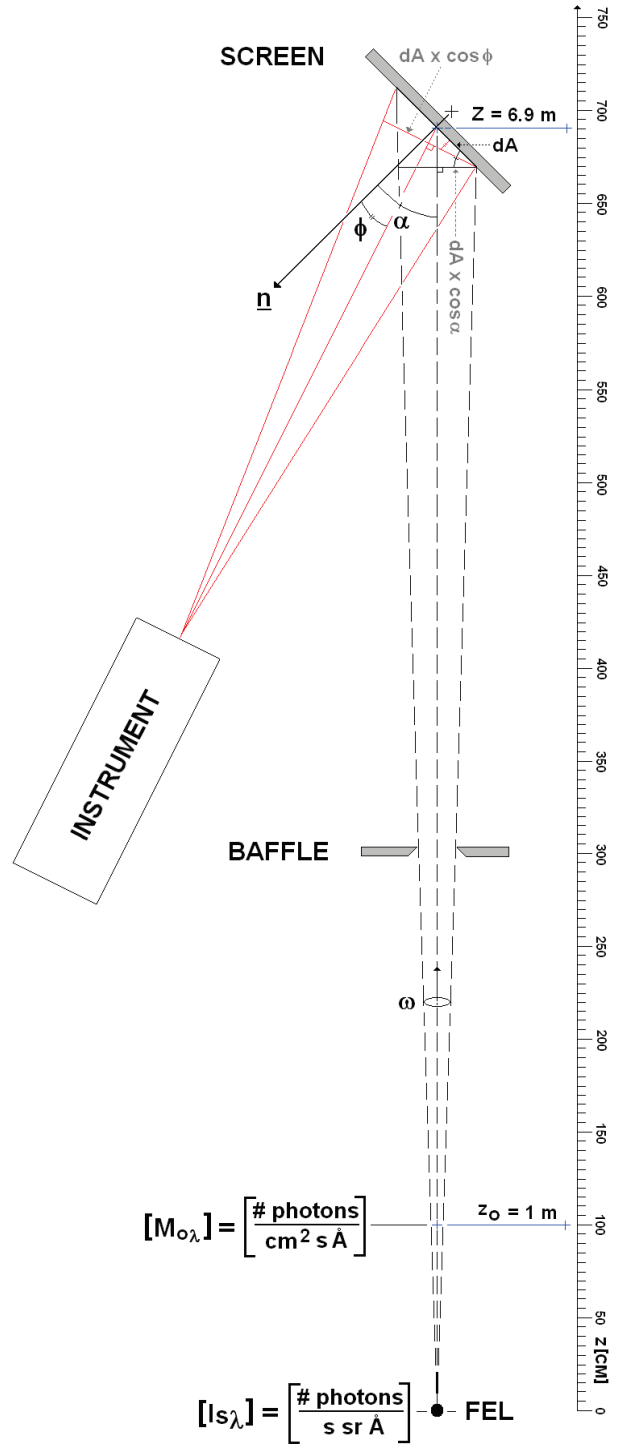


Figure 6. 2. Setup for absolute calibration of narrow field of view instruments.

The control room with the lamp power supply and a calibration spectrograph is located next to the lamp room and the screen room. A fiber bundle acts as entrance optics to the spectrograph, and it runs through a hole in the wall between the control room and the screen room. The distance between the lamp and the screen is controlled by mounting the lamp on a mobile table. Two rails are used to obtain smooth travel and constant horizontal centre distance parallel to the screen. The table is adjustable in height and a fixed laser is used to align the lamp with the vertical centre of the screen.

In this configuration, the maximum distance between screen and lamp is $z = 8.56\text{ m}$. The screen normal is pointing directly towards the lamp with $\alpha = 0$ degrees. The room lights are adjustable $+12\text{ VDC}$ lamps. When these lamps are turned off, they do not produce any low level background emissions as gas discharge tubes have a tendency to do.

The lamp power supply (ORIEL 68835 1KW) is coupled to a Light Intensity Controller (LIC). The intensity of the lamp is monitored and the current through the filament is adjusted to keep the intensity constant. The 1000 W lamp runs at 8.2 Ampere .

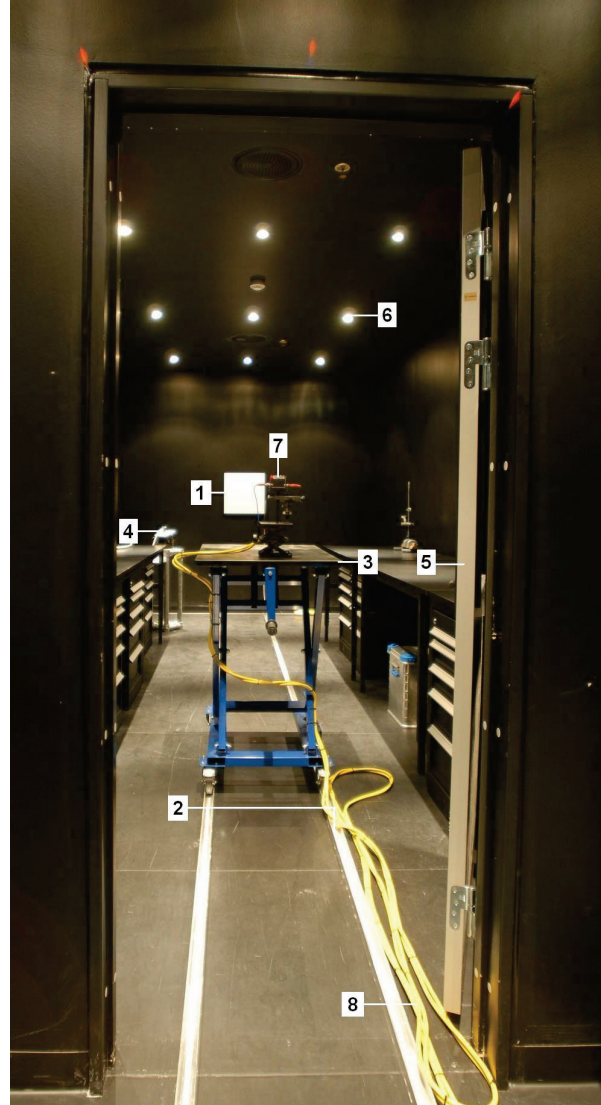


Figure 6.3. The calibration laboratory at UNIS: (1) $18 \times 18\text{ inch}^2$ Lambertian surface, (2) rails, (3) adjustable mobile table, (4) entrance fiber to spectrograph, (5) door with baffle, (6) room lights, (7) tungsten lamp, (8) power cable to lamp filament.

The main idea is to control the brightness of the screen only varying the distance z . See Eq. (6.15).

6.1.5 The lab spectrograph

The spectrograph is made by ORIEL. It uses a concave holographic grating (230 grooves/mm). The nominal spectral range is $4000\text{--}11000\text{ \AA}$. The detector is a 16-bit dynamic range thermoelectric cooled CCD from the company Hamamatsu (model INSTASPEC IV). Fig. 6.4 shows the optical diagram of the instrument. A field of view of 22° matches the fused silica fibre bundle used as entrance optics. The F-number is 2.1. The entrance slit is in the focal plane of the concave grating. The focal length varies from $129.4\text{--}132\text{ mm}$, depending on wavelength.

The diffracted light from the grating is focused and dispersed onto the exit plane. This is illustrated in Fig. 6.4, where the blue, green and red rays represent the start, centre and stop wavelength recorded by the CCD, respectively.

The bandpass is approximately 80 Å with a 100 μm wide entrance slit. The spectral resolving power of this instrument is moderate, but for our purpose it is enough. The spectrum of a tungsten lamp is smooth and continuous. There are no line structures that need to be resolved.

Wavelength calibration using spectral gas tube lamps with known emission lines must be carried out before any sensitivity calibration can take place. The most common way is to identify the pixel value associated with the a priori known wavelength of the emission line. At least three lines must be identified in order to minimize wavelength errors and check for non – linearity along the wavelength scale. The wavelength pixel relation is given as

$$\lambda \approx a_0 + a_1 \cdot p + a_2 \cdot p^2, \quad [\text{Å}] \quad (6.16)$$

where p is pixel value. The result of the procedure is the constants a_0 , a_1 and a_2 .

The FICS 77443 spectrograph was calibrated in wavelength using 3 different gas discharge mercury (Hg) lamps (see Fig. 6.5). Note that no order sorting filter is used to block out wavelength regions originating from higher spectral orders of the grating. The strong Hg emission line at wavelength 2537 Å is repeating itself 3 times at 5074 Å, 7611 Å and 10148 Å. The latter corresponds to spectral orders 2, 3, and 4, respectively. This effect is used in the calibration to obtain emission lines in the deep red part of the spectrum and up. Table 6.1 shows the result.

Coefficients	Constants
a_0	2.56026×10^{-3}
a_1	7.99624×10^0
a_2	1.95299×10^{-4}

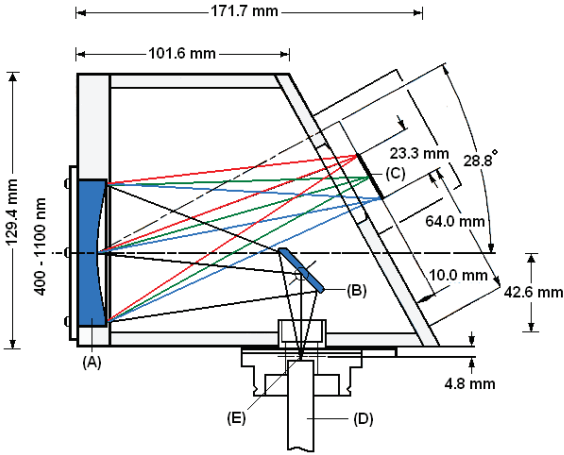


Figure 6.4. Fixed Imaging Compact Spectrograph (FICS SN 7743): (A) concave holographic grating, (B) flat mirror, (C) detector (CCD), (D) fiber bundle, (E) entrance slit.

Table 6.1. Wavelength calibration factors for the FICS 77443 spectrograph according to Eq. (6.16). The detector (CCD) has 256 x 1024 pixels. The spectral range is from 2560 Å to 10945 Å.

6.1.6 Transfer of lamp certificate

Our 1000 W tungsten lamp (ORIEL SN7-1275) is a traceable National Institute of Standards (NIST) source. This lamp is first used to calibrate the spectrograph in the wavelength region 4000 – 8000 Å. The calibration factor as a function of wavelength is then simply

$$\varepsilon_\lambda = \frac{B_\lambda}{u_\lambda^{1000W}}, \quad \left[\frac{\# \text{ photons}}{\text{cm}^2 \text{ s } \text{Å sr cts}} \right] \quad (6.17)$$

where u_λ^{1000W} is the count rate in units of counts per second [cts].

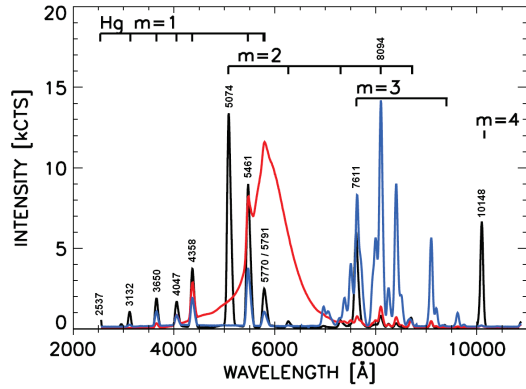


Figure 6.5. Wavelength calibration of FICS 44773 spectrograph. The spectrum plotted with the colour black is a low pressure mercury pen lamp (model MS-416 from Acton Research Corporation). The blue spectrum is from a mercury vapour tube supplied by Edmund Optics Ltd (SN K60-908). The red curve represents the spectrum of a battery powered fluorescent tube (OSRAM F6T5). Each mercury emission line is marked according to wavelength and spectral order m .

Next, the primary lamp is replaced by our secondary standard, a 200 W tungsten lamp. The lamp screen distance is 8.56 m and the exposure time is 160 $msec$ for both cases. This is important since the count rate as a function of exposure time is not necessarily linear. In addition, a cutoff filter is used in front of the fiber bundle to avoid overlapping spectral orders. The optical window is made of BK-7 and blocks out the UV part of the spectrum. Fig. 6.6 shows the results of the certification. Note that a certification above 8000 \AA requires an order blocking filter with cutoff wavelengths above 4000 \AA . A filter wheel will be installed to handle this issue in the near future.

As a temporally solution, the dotted black curve in Fig. 6.6 represents a functional fit to the 200 W secondary certificate based on a method developed by Saunders at NIST²⁰

$$M_{\lambda}^{200W} \approx \lambda^{-5} \times \exp(a - b\lambda^{-1}). \quad (6.18)$$

The equation has only two unknowns. In our case the solution becomes: $a = 73.9$ and $b = 52568$. The fit is obtained by using the irradiance wavelengths between 4000-8000 \AA .

6.1.7 Screen brightness control

According to Eq. (6.15) the brightness of the screen is adjustable as a function distance z . Another approach would be to vary the current through the lamp filament. A disadvantage of this technique is that the spectral shape changes with lamp current. It would also require a lamp certification for each current setting.

In order to save burning time or life time of the 1000 W tungsten lamp, the 200 W lamp is used to demonstrate the screen brightness control using only distance z as variable. The exposure time of the spectrograph is again kept constant at 160 $msec$.

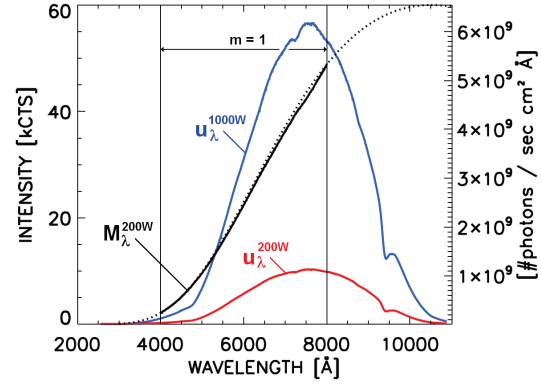


Figure 6.6. Spectra of Lambertian screen (SRT-99-180). The blue curve is raw counts in units of [kCTS] from the screen using the 1000 W tungsten lamp (ORIEL SN7-1275). Correspondingly, the red curve represents the 200 W secondary tungsten lamp. Exposure time is 160 $msec$. Screen lamp distance is $z = 8.56$ m . The 1000 W lamp runs on 109.6 V at 8.20 A . The 200 W lamp has a filament current of 6.50 A at a voltage of 31.7 V . The screen angle $\alpha = 0^\circ$. The black curves are the brightness of the screen in absolute units [$\#photons / cm^2 s \text{\AA}$] using the 200 W lamp. The valid wavelength region of certification is marked for spectral order $m=1$.

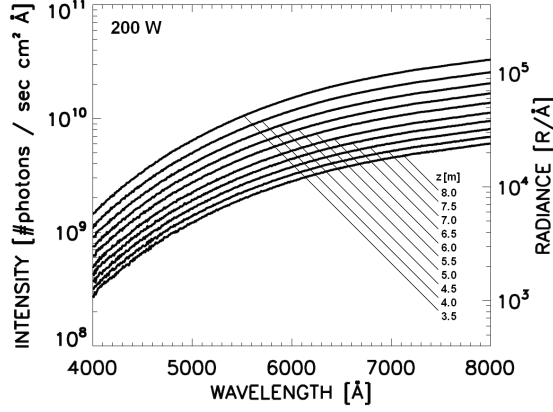


Figure 6.7. Calibrated spectra of Lambertian screen (SRT-99-180) as a function of screen to lamp distance z . The source is the 200 W tungsten lamp (FRED SN01).

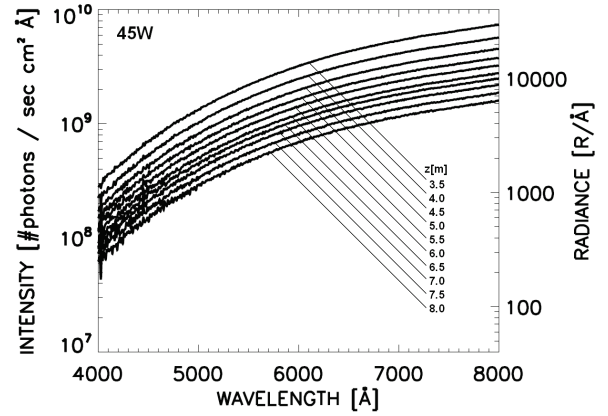


Figure 6.8. Calibrated spectra of Lambertian screen (SRT-99-180) as a function of screen to lamp distance z . The source is a 45 W tungsten lamp (ORIEL SN7-1633).

The mobile table is moved towards the screen in incremental steps of 0.5 m . As seen in Fig. 6.7 the spectral shape remains the same and the intensity rises with the inverse of the distance squared.

The above procedure was repeated using a 45 W tungsten lamp (ORIEL SN7-1633). Fig. 8 shows the results. Intensities are now down by a factor close to 5. Note that the signal to noise, especially in the blue part of the spectrum, may be improved by simply rising the exposure time of the spectrograph. An exposure time of 4000 instead of 160 $msec$ at a distance of 8 m , generate raw spectra that span the whole dynamic range of the detector (16-bit).

Note that the left ordinates of Fig. 6.7 and 6.8. are given in units of Rayleigh per Ångström [$R/\text{Å}$]. The generalized definition is ^{21, 22}

$$1R \equiv (1/4\pi) \times 10^6 \# \text{photons cm}^{-2} \text{sec}^{-1} \text{sr}^{-1}, \quad (6.19)$$

and it applies directly to the radiance B_λ . In Eq. (6.15), a multiplication factor of $(4\pi / 10^6)$ is used to convert the number flux of $\# \text{photons cm}^{-2} \text{sec}^{-1} \text{sr}^{-1} \text{Å}^{-1}$ to $R/\text{Å}$.

6.1.8 Experimental uncertainty

The 1000 W tungsten lamp supplied by Oriel is provided by a calibration uncertainty of $\pm 3 \%$ in the wavelength range 4000 – 8000 Å . The life time of the lamp certificate is 50 hours. After this, the lamp should be recalibrated. Oriel reports drift rates in the range 1 to 2 % per 100 operating hours over the wavelength range 4000 to 8000 Å , respectively.

Furthermore, our secondary transfer procedure will introduce errors ^{23, 24} due to distance (0.14 %), lamp orientation and alignment (0.3 %), spectrograph stability (0.3 %), measurement repeatability (0.15 %), spectrograph nonlinearity (0.2 %) and laboratory stray light (0.2 %) ²⁵.

Based on the above numbers our total calibration transfer uncertainty is estimated to be below $\pm 4 \%$.

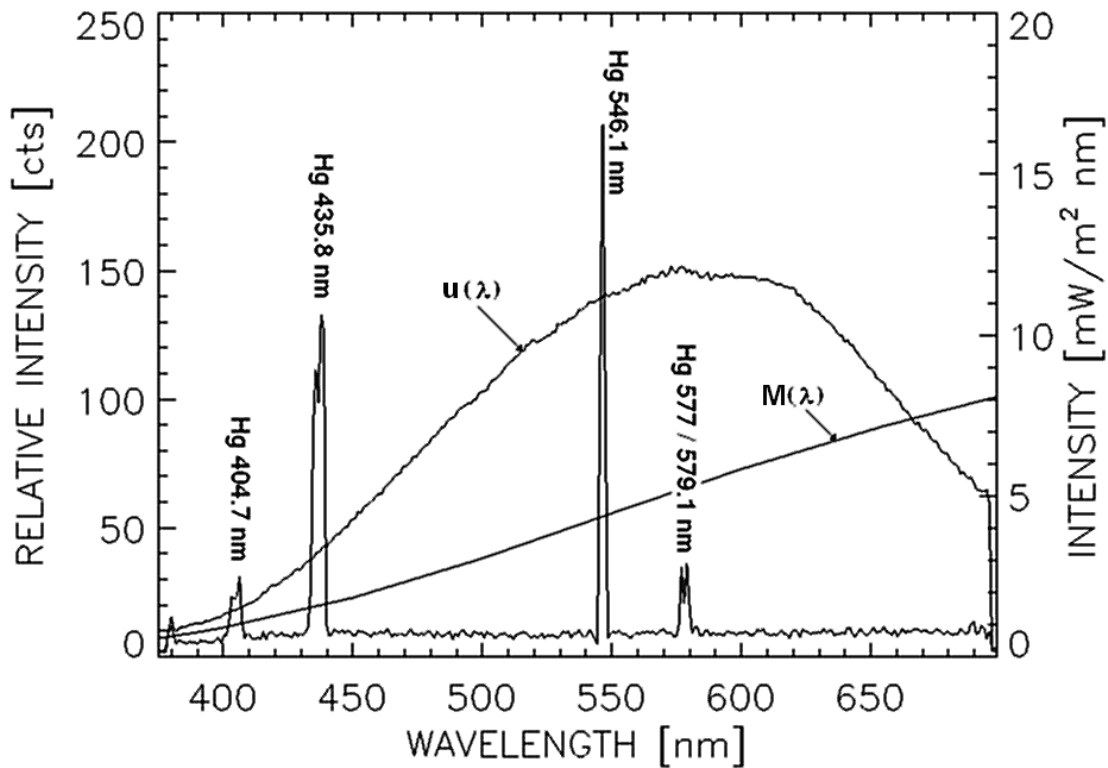


Figure 6.9. Result of both wavelength and absolute calibration of the DRONESPEX hyperspectral imager. The wavelength calibration was obtained by the use of a mercury vapor gas (Hg) lamp. The mercury emission lines at 404.7, 435.8, 546.1 and 577 / 579.1 nm are marked. $u(\lambda)$ is the spectrum of the diffuse re-emitting screen illuminated by the absolute calibration lamp (Tungsten). Note that u is the average slit response given in raw counts [CTS]. $M(\lambda)$ represents the absolute certificate spectrum for the 1000W Tungsten lamp in units [$mW m^{-2} nm^{-1}$].

6.1.9 DRONESPEX calibration

Fig. 6.9 shows the result of the calibration for our hyperspectral instrument (Chapter 5). It is clear from Fig. 6.9 that the instrument did not detect many counts in the blue range of the spectrum due to the low output of the calibration lamp. A better result would be to use the sun itself as a calibration source. The total radiation on a clear day will produce a higher throughput in the blue part – optimizing the instrumental count rate to be more flat across the spectrum. Even an overcast day would work if the cloud cover does not change during calibration. The setup requires that we know the intensity of the screen in absolute units. This can be obtained using a second instrument that is already calibrated in the laboratory by the procedure outlined above. The second instrument does not need to have imaging or hyper spectral capability – a regular spectrograph with a line array detector sensitive in the blue will be sufficient.

6.1.10 Summary spectral calibration

The calibration laboratory at UNIS has been constructed to calibrate narrow field of view instruments with variable source brightness control.

Two tungsten lamps with power of 200 W and 45 W are certified in the visible part of the spectrum (4000 – 8000 \AA) within an uncertainty of $\pm 4\%$. By varying the screen to lamp distance from 8 m down to 3.5 m , screen intensities in the range 200 – 132 $k R/\text{\AA}$ were detected.

The outlined calibration procedure will be repeated to include extended wavelength coverage by using proper cut off filters to block out higher overlapping spectral orders of the FICS 7743 spectrograph. In addition, we aim to use a tungsten lamp of lower power (12 – 20 W) to cover intensities below 200 $R/\text{\AA}$.

6.2 Bandpass filter instrument calibration

This section shows how to calibrate a bandpass filter instrument. One example instrument is the photometer. It consists of a photomultiplier tube as detector, a focusing lens and a narrow bandpass filter.

The experimental setup for a typical bandpass filter instrument sensitivity calibration is sketched in Fig. 6.10.

The type of filters used depends on center wavelength λ_c , bandpass BP and transmission T_λ . The most traditional one is the Fabry-Perot filter design, where parallel transparent glass plates create interference due to multiple reflections between the plates. Panel (B) of Fig. 6.10 shows a typical transmission curve for these filters.

The raw data counts of the detector D is given as an integral over wavelength

$$u = \int B_\lambda \cdot S_\lambda d\lambda, \quad [cts] \quad (6.20)$$

where S_λ is defined as the spectral responsivity. B_λ is given by Eq. (6.15).

If we assume that B_λ , lens transmission and detector sensitivity varies slowly in the wavelength interval $\Delta\lambda$, then the spectral responsivity of the instrument may be expressed as

$$S_\lambda = \varepsilon \cdot T_\lambda \cdot [cts R^{-1}] \quad (6.21)$$

ε is now an instrumental constant. It then follows that

$$u = \int B_\lambda \cdot S_\lambda d\lambda = B_{\lambda_c} \int \varepsilon \cdot T_\lambda d\lambda = B_{\lambda_c} \cdot \varepsilon \int T_\lambda d\lambda = \varepsilon \cdot B_{\lambda_c} \cdot A, \quad (6.22)$$

where A is the area of the filter transmission curve. The calibration factor is then simply

$$\varepsilon = \frac{u}{B_{\lambda_c} \cdot A}. \quad (6.23)$$

Lets us assume that T_λ is narrow and triangular in shape. Then A becomes

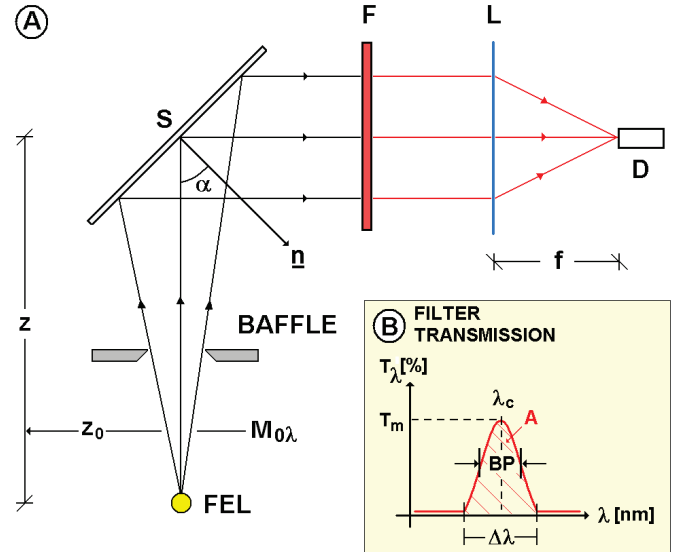


Figure 6.10. Panel (A): Filter instrument sensitivity calibration setup: F bandpass filter, L focusing lens, D detector, S Lambertian screen, FEL standard Tungsten light source. Panel (B): Sketch of a typical transmission curve for a bandpass interference filter.

$$A = \int T_{\lambda} d\lambda \approx T_m \cdot BP. \quad (6.24)$$

T_m is the peak transmission of the filter at $\lambda = \lambda_c$.

Furthermore, we define the intensity of a discrete auroral emission line at wavelength λ_c to be

$$J = J_0 \cdot \delta(\lambda - \lambda_c),$$

where δ is the Kronecker delta and J_0 is the absolute intensity. The number of auroral raw data counts is then

$$u_a = \int J \cdot S_{\lambda} d\lambda = \int J_0 \cdot \delta(\lambda - \lambda_c) \cdot \varepsilon \cdot T_{\lambda} d\lambda = \varepsilon \cdot J_0 \cdot \int T_{\lambda} \cdot \delta(\lambda - \lambda_c) d\lambda = \varepsilon \cdot J_0 \cdot T_m \quad (6.25)$$

Finally, from Eqs. (6.23) and (6.24) we obtain

$$\begin{aligned} J_0 &= \frac{u_a}{\varepsilon \cdot T_m} = \left(\frac{u_a}{u} \right) \times \left(\frac{A}{T_m} \right) \times B_{\lambda_c} \quad [R] \\ &= \left(\frac{u_a}{u} \right) \times B_{\lambda_c} \times BP \end{aligned} \quad (6.26)$$

Note that Eq. (6.26) only applies when we assume that the transmission profile of the filter is narrow and triangular. This is a rather crude assumption, especially for medium to broad bandpass filters.

6.3 Color camera calibration

A method to obtain the average spectral pixel responsivity and the quantum efficiency of Digital Single Lens Reflex (DSLR) cameras is outlined^{33,34}. As examples, two semi-professional cameras, the Nikon D300 and the Canon 40D, are evaluated.

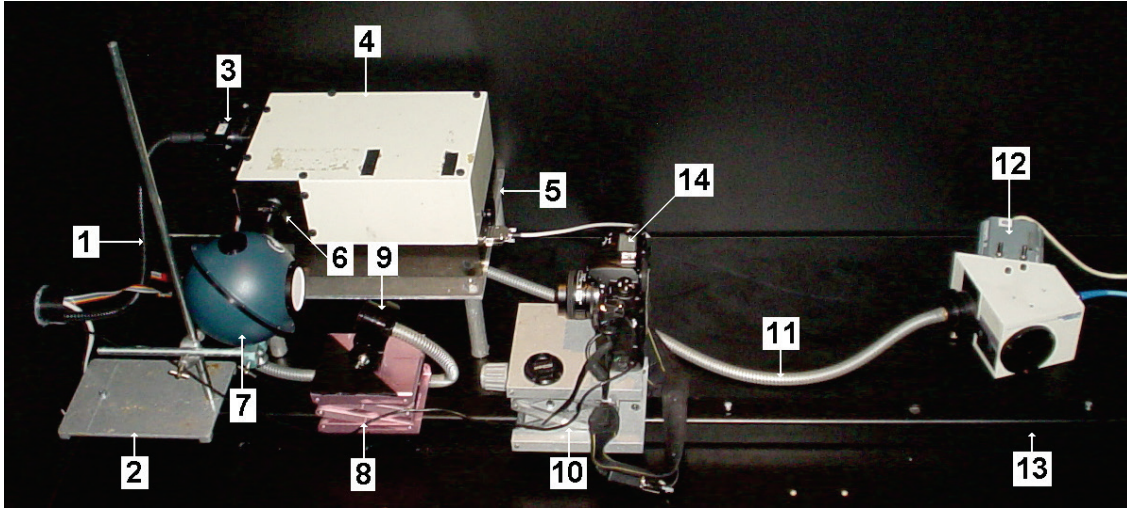


Figure 6.11. Experimental setup: (1) fiber bundle from lamp located in neighboring room, (2) mount / stand for integrating sphere, (3) order sorting filter wheel in front of entrance slit, (4) Jobin Yvon HR320 monochromator, (5) table, (6) exit slit plane, (7) Edmund Scientific integrating sphere, (8) laboratory lift table, (9) fiber bundle holder, (10) camera table, (11) fiber bundle used as input to spectrograph, (12) Oriel FICS 7743 spectrograph, (13) optical mount rail, and (14) DSLR camera with normal 50 mm f/1.4 objective.

6.3.1 Short Background

The main motivation for this work is to develop a method to obtain the spectral responsivity or the quantum efficiency of each pixel in a DSLR camera. These cameras have become valuable tools for studies of the night sky including phenomena such as aurora and airglow. It is therefore essential to establish a method to quantify the light throughput of the cameras. Even though our focus is on the aurora, the results should be of interest to other fields of science such as astronomy, remote sensing and industrial processing as well.

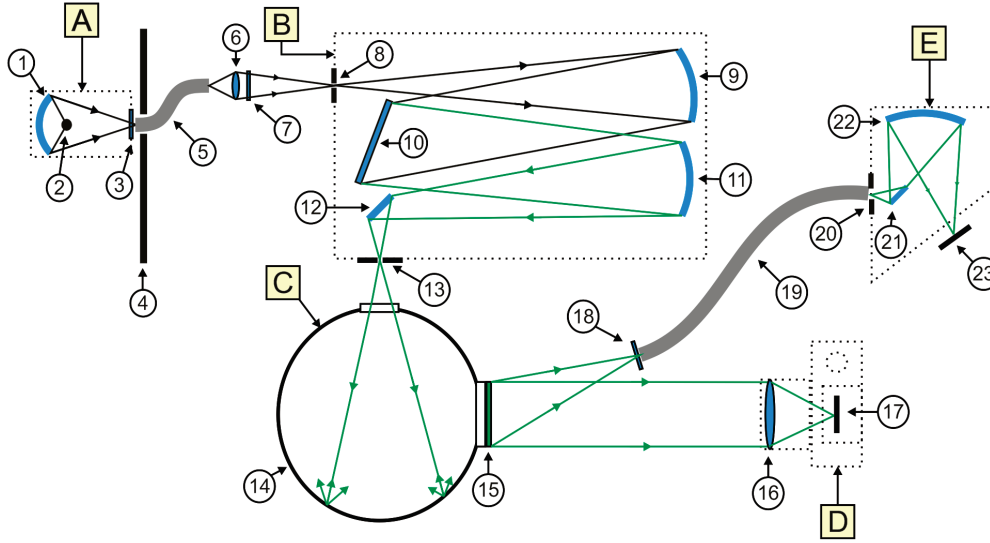


Figure 6.12. Optical diagram. [A] Leica 150W fiber illuminator: (1) mirror, (2) Tungsten filament, (3) heat filter, (4) blocking wall, and (5) fiber bundle. [B] Jobin Yvon HR320 Monochromator: (6) f-matching lens, (7) order sorting filter, (8) entrance slit, (9) collimator mirror, (10) plane reflective grating, (11) focusing mirror, (12) flat surface folding mirror, and (13) exit slit. [C] Edmund Scientific General purpose 6 inch diameter integrating sphere: (14) sphere, and (15) transmitting diffuser (Teflon). [D] DSLR camera: (16) 50 mm normal f/1.4 objective, and (17) CMOS / CCD detector. [E] Oriel FICS 7743 spectrograph: (18) order sorting filter, (19) fiber bundle, (20) entrance slit, (21) folding mirror, (22) concave grating, and (23) CCD detector.

6.3.2 Experimental setup

Figs. 6.11 and 6.12 show the experimental setup and the optical diagram of the system, respectively. The main components are a fiber illuminator (Leica 150W) that is connected to the entrance slit of a monochromator (Jobin Yvon HR320). The diffracted light at the exit slit of the monochromator is then fed into the integrating sphere. The output of the sphere is the target for both the DSLR cameras and the intensity calibrated FICS spectrograph (see section 6.1.5).

The sphere, from Edmund Optics, is designed to integrate radiant light fluxes. The sphere is 6 inches in diameter. Its interior walls are coated with Spectralon that has a diffuse reflectance factor of 0.98 throughout the visible part of the spectrum. The monochromatic light that enters the 1 inch diameter input port of the sphere is scattered multiple times before it exits the 2.5 inch diameter output port.

In addition, a transmitting diffuser is used at the output port of the integrating sphere to diffuse the light even more (see Fig. 6.12). The diffuser is 0.5 mm thick and made of Teflon (opal). This was found necessary due to the fact that the first region of illumination in the sphere, caused by reflected light directly from the exit slit of the monochromator, is easily seen from the output port. This illuminated region is rather large compared to the diameter of the sphere. The effect was clearly seen even at relatively small off-axis view angles to the output port.

Finally, the net result is a uniform illuminated surface that is comparable to the Lambertian surface that is used for calibrating narrow field of view instruments such as our spectrograph. The field of view of the spectrograph is fully illuminated by an area that is uniform in intensity, and equal to the corresponding area that fills the field of view of each selected pixel of the cameras. As long as the field of view of the pixels and the spectrograph are filled neither the look angle nor the distance to the diffuser matters. The changing sizes of the pixel fields of view at the diffuser surface compensates exactly for both the changing distances and angles.

Both the Nikon D300 and the Canon 40D cameras are operated in manual mode with sensitivity set at ISO 1600. The Nikon D300 uses a Nikon 50mm f/1.4 AF-D lens, while the Canon 40D uses the Canon EF 50mm f/1.4 USM lens. The lenses are operated at maximum aperture (f-value set to 1.4). Exposure times between 3 and 4 seconds were chosen to avoid overexposures.

6.3.3 The spectral pixel responsivity

The assembled wavelength tunable system is designed for the visible part of the spectrum (4000-7000 Å), producing monochromatic lines with a bandpass of ~12 Å. The calibrated FICS spectrograph measures the intensity of the integrating sphere output in units of $R/\text{Å}$. The bandpass of the spectrograph is $\Delta\lambda \approx 100 \text{ Å}$. As result, the spectral responsivity for each camera is calculated using all 31 spectra and images. The equation of observations is

$$\hat{u}^{(k)} = B \cdot \hat{S}^{(k)} \cdot \Delta\lambda, \quad [cts] \quad (6.27)$$

where the subscript $k \in [r, g, b]$ labels the red, green and blue channels of the

transparent Colour Filter Mosaic filter (CFM) in front of each pixel of the sensor, respectively. Eq. (6.27) is the same as Eq. (6.20) expressed in vector form. The vector $\hat{u}^{(k)}$ contains 31 numbers of averaged pixel raw counts per second for each colour channel, k . The matrix B is given as

$$B = [B_1 B_2 B_3 \cdots B_{31}]^T. \quad [R/\text{Å}] \quad (6.28)$$

Each vector B_i , $i \in [1 \dots 31]$ contains the spectrum as measured by the FICS spectrograph.

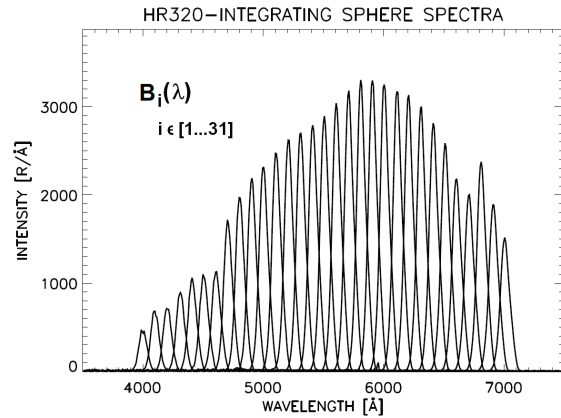


Figure 6.13. Sphere source functions. $B_i(\lambda)$ is the set of observations that consists of 31 spectra from the monochromator (HR320) illuminating the 6 inch diameter integrating sphere from Edmund Optics.

The dimension of the spectra is reduced to 31 by re-sampling at each center wavelength setting of the monochromator. The spectral responsivity $\hat{S}^{(k)}$ is now found by solving Eq. (6.27) by Singular Value Decomposition (SVD).

6.3.4 The quantum efficiency

Another way to characterize the spectral sensitivity of a camera is to calculate the Quantum Efficiency (QE). It is defined as the fraction of photons that will generate electrons detectable by the photo-reactive sensors. In our case with monochromatic source functions in units of $R/\text{\AA}$, the QE may be expressed as

$$QE_i^{(k)} \approx \left[\frac{4\pi \cdot u_i^{(k)} \cdot \Delta t \cdot g}{10^6 B_i \cdot \Delta \lambda \cdot \Delta A} \right] \times 100, \quad [\%] \quad (6.29)$$

where ΔA is the pixel area in units of cm^2 , Δt is the exposure time in seconds and g is the gain of the detector defined as the conversion factor between the number of electrons and raw counts per pixel. A simple procedure of how to measure the gain is given in Handbook of Astronomical Image Processing³⁵. At ISO 1600 the gain is 0.775 and 0.675 electrons per 12-bit data count for the 40D and the D300, respectively.

6.3.5 Calibration results

Our library of calibrated spectra is shown in Fig. 6.13. The intensity of the output port of the integrating sphere rises from $\sim 500 R/\text{\AA}$ at 4000\AA to a maximum of $\sim 3300 R/\text{\AA}$ at 5800\AA , reducing to $\sim 1500 R/\text{\AA}$ at 7000\AA . The dips at 4600\AA and 6700\AA are due to drops in the grating efficiency of the monochromator.

Fig. 6.14 shows the Nikon D300 green channel raw counts per second when the sphere is illuminated with light at 5569\AA center wavelength. Note that dark frame subtraction is carried out to reduce the background level. The resulting image is shaped like a steep mountain cliff with a sharp leveled circular plateau. The height of the plateau is used as the average pixel raw count response. The plateau height is calculated by selecting raw count values within a 200×200 pixel square area in the center of the image. Using the notation in Section 6.3.3, the plateau height is $u^{(G)} = 964 \text{ cts/s}$. The corresponding standard deviation is $2\sigma = \pm 30 \text{ cts/s}$, which in percentage terms is close to 3% variation across the plateau. The 2σ deviation is compatible with Poisson statistics where noise is defined as the square root of the counts.

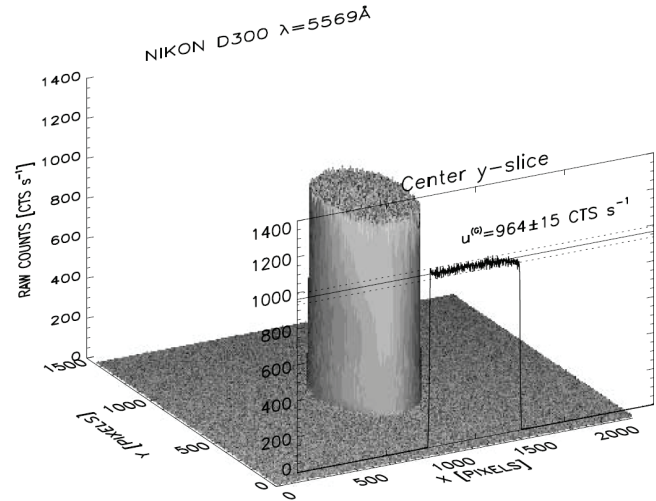


Figure 6.14. Raw counts from the Nikon D300. Exposure time is 3 seconds at ISO 1600 with a 50 mm normal objective fixed at f/1.4. The shaded surface represents the green channel 12-bits raw cts per second. The source is a monochromatic illuminated sphere at center wavelength 5569\AA . The center y-axis slice of the surface is shown together with the calculated height of the plateau (solid horizontal line) and the 2σ standard deviation (dotted lines).

The calculated values are visualized in Fig. 6.14 by plotting the center y-axis slice of the mountain cliff. The result looks like a square pulse with ripple noise on the top that is within the calculated deviation. The output of the sphere is, in other words, sufficiently uniform to assume that the above 200x200 pixel average is the same for all pixels.

The next step is to repeat the procedure to obtain the average raw counts per second as a function of wavelength for each color channel of the cameras. The first thing we noticed when handling the raw data from the Canon 40D is that the counts are fixed at 14-bits resolution. The range is, in other words, from 0 to 16383 *cts*. The Nikon D300 has a 12-bit range (0 – 4095 *cts*). As a consequence, the 40D counts are scaled down by a factor of (1/4) in order to be comparable to the D300. The D300 was first operated at 3 seconds exposures with no overexposures occurring as we changed the wavelength of the monochromator. At 4 seconds the green channel became overexposed. On the other hand, the Canon 40D did not overexpose at 4 seconds exposure time. It also turned out that we could not set the exposure time of the 40D to 3 seconds. Only 2.5, 3.2 and 4 second intervals were possible. The procedure was therefore repeated to make sure that we obtain the same cts per second for all intervals.

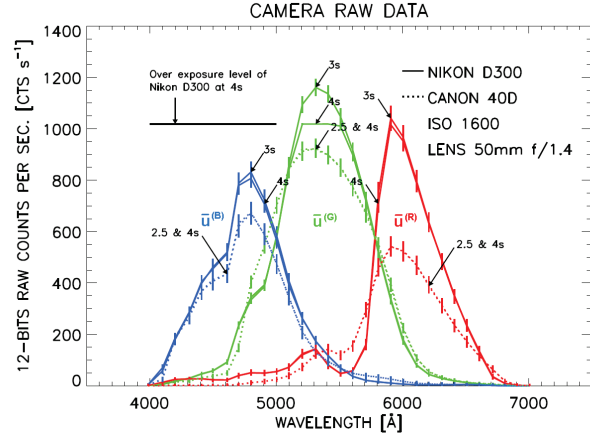


Figure 6.15. Camera raw data. Solid lines are the raw counts per second from the Nikon D300 camera for each color channel (Red, Green and Blue). The dotted lines are the corresponding data from the Canon 40D camera. Each curve is labeled with exposure time settings of the cameras. Both cameras were operated with identical settings using normal objective lenses (50mm f/1.4) at ISO 1600. The error bars represent the 2σ standard deviation of the count rates.

Fig. 6.15 shows the resulting raw counts per second obtained with variable exposure intervals for both cameras. It is hard to see any difference between the curves as a function of exposure time for each camera alone, especially for the 40D. The only difference appears when the D300 green channel is overexposed at 4 seconds. The effect is seen comparing the 3 and 4 second exposures. The peak of the 4 second exposure looks like it has been cut off compared to the 3 second one in the 5200 to 5500 Å wavelength region. Also notice that the error bars become zero in this region. This is expected since the error bars are defined as the 2σ standard deviation of the count rates. The D300 blue and red channel count rates are almost identical, with deviations that are well within the error bars. It is therefore reasonable to assume that the raw count rates are constant as a function of exposure time in the interval ~ (3-4) seconds for both cameras.

The count rate profiles of the cameras shown in Fig. 6.15 are quite similar in shape, especially the blue and red channels. However, the 40D green channel profile is slightly more symmetric than the corresponding D300 profile. This is due to the fact that the 40D green counts are little bit higher than the D300 green counts in the 4800 to 5100 Å wavelength region. The same is seen in the 5400 to 5800 Å wavelength region when comparing the red channels.

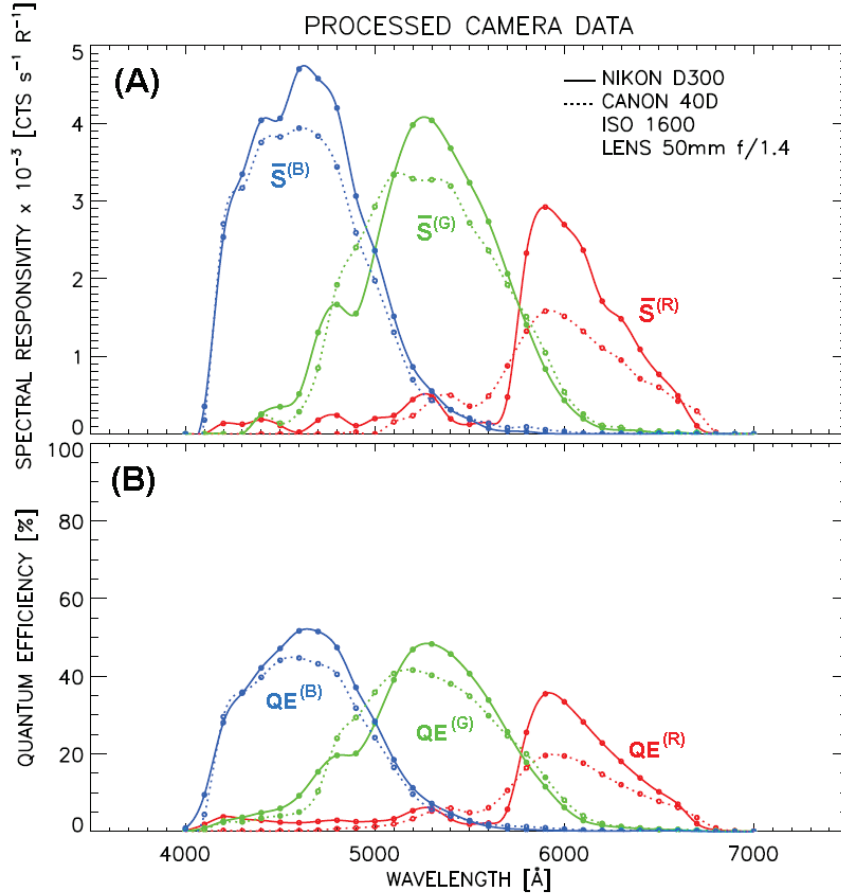


Figure 6.16. Processed camera data. Panel (A): Solid lines are the spectral responsivity of the Nikon D300 camera for each color channel (Red, Green and Blue). The dotted lines are for the Canon 40D camera. Panel (B) shows the corresponding calculated quantum efficiency. Both cameras were operated with identical settings using normal objective lenses (50mm f/1.4) at ISO 1600.

Both cameras have peak count rates in the green channels. However, the D40 has a higher blue than red peak count rate, while this is vice versa for the D300. What this means, in terms of difference in color balance between the cameras, depends on the spectral responsivity or the quantum efficiency. The raw count rate profiles alone are not sufficient in order to conclude on this issue.

If we define the width of the profile for each color channel to be the wavelength region where it has equal to, or greater than, half of its maximum count rate, then it is clear that the D40 has wider profiles than the D300 for all color channels. The D40 profile is $\sim 250 \text{ Å}$ wider than the D300 in the green channel. The corresponding blue and red differences in widths are 70 and 160 Å, respectively.

The main difference between the cameras becomes evident in the level of the count rates. The Nikon D300 count rates are generally higher when compared to the Canon 40D, except for the wavelength regions mentioned above. By integration we find that the D300's blue, green and red channels have ~ 16 , 8 and 50 % higher count rates compared to the D40, respectively

The spectral responsivity and the quantum efficiency may now be solved according to equations (6.27) and (6.29), respectively. Based on the above findings, the 3 second exposure by the D300 and the 4 second exposure by the 40D camera are chosen to represent the raw count rates in the calculations. The net result is shown in Fig. 6.16 for both cameras.

The calculated spectral responsivity and the quantum efficiency curves are compatible in shape and amplitude. Note that the quantum efficiency calculation is a more direct and robust method since it does not depend on the Singular Value Decomposition. It is now clear that both cameras have their peak sensitivity in the blue and minima in the red channels. The color balance between the channels is, in other words, the same for both cameras. However, as expected from the level of the count rate profiles, the Nikon D300 is the most sensitive camera. The D300 has a peak quantum efficiency of 50% and spectral responsivity of $4.3 \times 10^{-3} \text{ cts s}^{-1} R^{-1}$ in the blue channel at 4600 \AA . The green channel quantum efficiency peaks at 48% with $3.9 \times 10^{-3} \text{ cts s}^{-1} R^{-1}$ in spectral responsivity at 5300 \AA , while the red channel peak is lowest at 35% and $2.8 \times 10^{-3} \text{ cts s}^{-1} R^{-1}$ at 5900 \AA . Again, by integration we find that the D300's blue, green and red channels have ~ 17 , 9 and 54% higher spectral responsivity compared to the D40, respectively. The corresponding difference in quantum efficiency is 11, 3 and 47%.

The above is a surprising result. We double checked our experimental setup to make sure that we actually used the same settings on both cameras. Also note that, the total difference between the two sets of source functions used to obtain the spectral responsivity and quantum efficiency only varied by $\pm 17 R/\text{\AA}$. The difference is so small that it is hard to see it if we overlay the curves of the 40D source functions in Fig. 6.13.

One factor that could explain the discrepancy, especially in the red, could be the difference in the spectral transmission of the lenses. The Nikon 50mm f/1.4 AF-D and the Canon EF 50mm f/1.4 USM lenses are almost identical with the same number of optical elements in the lens construction. Fig. 6.17 shows the spectral transmission profiles using the FICS spectrograph as the detector and a flat Lambertian surface as the target reference. The surface was illuminated by a Tungsten lamp. The lenses were mounted in front of the spectrograph's entrance fiber bundle. The setup is identical to our narrow field of view intensity calibration procedure

The shape of the transmission profiles is more or less equal for both lenses. There is a gradually increase from $\sim 80\%$ at 4000 \AA up to a more stable region above $\sim 90\%$ from 4500 to 6500 \AA . In the deep red part of the spectrum ($\lambda > 6500 \text{ \AA}$), the transmission factors start to decrease more rapidly to a level below ~ 50 to 60% at $10\,000 \text{ \AA}$.

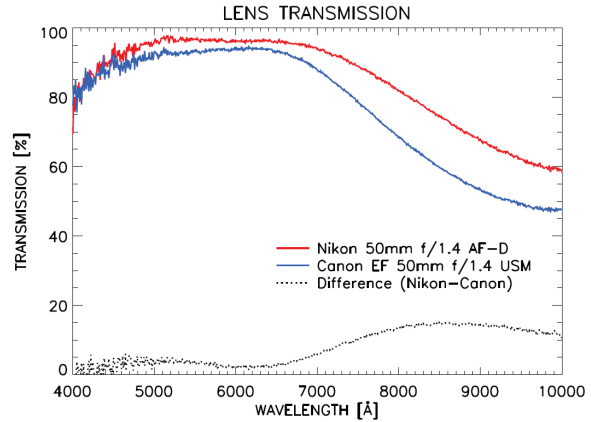


Figure 6.17. Normal objective spectral lens transmissions for the Nikon 50mm f/1.4 AF-D (red solid line) and the Canon EF 50mm f/1.4 USM (blue solid line). The dotted black line is the difference in transmission between the Nikon and Canon lens, respectively.

The Nikon lens is overall more effective than the Canon lens throughout the visible region (4000 – 7000 Å) and the near infra-red (7000 - 10000 Å) by a factor of ~5 and 15%, respectively.

If we take into account the transmission of the lenses, the difference between the cameras in spectral responsivity and quantum efficiency is only changed by a few percent. The D300's blue, green and red channels now have ~12, 5 and 50% higher spectral responsivity compared to the D40, respectively. The updated differences in quantum efficiency become ~7% for the blue, 0% for the green and 45% for the red channels.

The above result indicates that the main difference between these two cameras is in their detectors. Both cameras use a CMOS (Complementary Metal Oxide Semiconductor) sensor and a Color Filter Mosaic (CFM) to separate the colors. The main difference in spectral responsivity is found to be in the red channels, and could well be related to the transmission of the red elements in the CFM, the infra-red filter or in the semiconductors used. This is an interesting topic, but it is beyond the scope of this paper.

6.3.6. Lesson learned

A digital camera image could contain more information than just relative scaled intensities or color coded pixel values that have little physical meaning in terms of brightness on a quantitative scale. The recent improvements in both sensitivity and dynamic range enable the DSLR camera to be used as an intensity tool as well. But the lesson learned from this exercise is that there seems to be a lack of a common standard for camera sensitivity given by the manufacturers. As shown above for Nikon and Canon, the ISO (International Organization for Standardization) values, originally defined as the speed of photographic film, may not be the optimum parameter representing the sensitivity of a digital sensor. The spectral responsivity or the quantum efficiency is the parameter that should be used in the future to characterize the sensitivity of a camera. We hope that the manufacturers can provide this information in future. It would increase the usage and potential of these fantastic devices.

7. ATTITUDE CORRECTIONS

This chapter presents the theory of Least Square adjustment and Kalman filtering. The theory is applied to a simple Inertial Measurement Unit (IMU) mounted to a dynamic moving system. The aim is to compile software routines that enable us to estimate the correct attitude for our airborne spectral imagers and cameras.

7.1 Introduction

In contrast to a static frame of reference, a body in a dynamic or moving system is influenced by random accelerations / forces from nearly all directions. The body's attitude obtained by a static sensor will as a consequence produce errors. A pendulum mounted inside a car can be used as an example to illustrate this. If the car is at rest or moving with constant speed, the pendulum's angle with the vertical represents a good measure of the car's tilt angle. On the other hand, when the car is accelerated or decelerated the pendulum will start to swing as a response to the forces acted up on it. As a result, large tilt errors occur.

In order to obtain the correct attitude it is necessary to measure both the acceleration and the turn rate. Typical sensors are 3 axis solid-state accelerometers combined with turn rate gyros - often called Inertial Measurements Units (IMU). These sensors are rather noisy and tend to drift in time, especially the turn rate sensors. Therefore, a filter to estimate the correct attitude is needed. There are numerous techniques that can be used, but Kalman filtering is the only one presented here.

In order to understand the Kalman filter, we first introduce the theory of Least Square adjustment. The theory is based on work done by Hofmann-Wellenhof and Lichtenegger²⁶. The Kalman filter sequence is adopted from Brown and Hwang²⁷.

7.2 Least squares adjustment

7.2.1 Standard adjustment

Least square adjustment is based on equations where the observations are expressed as a function of unknown parameters. In matrix notation the linear observation model is expressed as

$$\underline{l} = \underline{A} \underline{x}, \quad (7.1)$$

where \underline{l} is the vector of observations (measurements), \underline{A} the design matrix and \underline{x} the vector of unknowns. In case of nonlinear functions, a Taylor series expansion is usually performed. The series is then truncated after the second term in order to obtain a linear function. Furthermore, we introduce the covariance matrix of the measurements

$$\underline{Q}_l \cong E[\underline{l} \cdot \underline{l}^T] = \begin{bmatrix} \sigma_1^2 & \rho_{12} \cdot \sigma_1 \cdot \sigma_2 & \cdots & \cdots & \rho_{1n} \cdot \sigma_1 \cdot \sigma_n \\ \rho_{12} \cdot \sigma_2 \cdot \sigma_1 & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho_{1n} \cdot \sigma_n \cdot \sigma_1 & \cdots & \cdots & \cdots & \sigma_n^2 \end{bmatrix}. \quad (7.2)$$

Here σ_i^2 is the squared variance of the measured l_i variable, and ρ_{ij} is the correlation factor between l_i and l_j . Moreover we have $\rho_{ij} = \rho_{ji}$. In our cases there is no correlation between the sensors, thus $\rho_{ij} = 0$ if $i \neq j$. $E[.]$ is the mathematic hope or expectation value. The inverse of the covariance matrix

$$\underline{P} = \underline{Q}_l^{-1} \quad (7.3)$$

is called the weight matrix. If we assume n observations and u unknown parameters, the design matrix \underline{A} consists of n rows and u columns. The system is redundant when $n > u$. In most of the cases, $n = u$. The inconsistency is due to observational errors or noise. In order to solve the model and assure consistency, the noise vector \underline{n} is added to the vector of observations

$$\underline{l} = \underline{A} \underline{x} + \underline{n}. \quad (7.4)$$

The solution of (7.4) becomes unique and consistent if we apply the least square principle, which state that the weighted sum of the squared residuals is minimized²⁸.

In terms of our notation the least square principle is given by

$$\frac{\partial [\underline{n}^T \underline{P} \underline{n}]}{\partial \underline{x}} = 0 \quad (7.5)$$

If we apply this minimum principle to the above observational model, we obtain first the following equation

$$\underline{n}^T \underline{P} \underline{n} = (\underline{l} - \underline{A} \underline{x})^T \underline{P} (\underline{l} - \underline{A} \underline{x}) \quad (7.6)$$

or

$$\underline{n}^T \underline{P} \underline{n} = (\underline{A} \underline{x})^T \underline{P} \underline{A} \underline{x} - \underline{x}^T \underline{A}^T \underline{P} \underline{l} - \underline{l}^T \underline{P} \underline{A} \underline{x} + \underline{l}^T \underline{P} \underline{l}. \quad (7.7)$$

Secondly after derivation (see Appendix E)

$$\frac{\partial [\underline{n}^T \underline{P} \underline{n}]}{\partial \underline{x}} = \underline{A}^T \underline{P} \underline{A} \underline{x} + \underline{A}^T \underline{P}^T \underline{A} \underline{x} - \underline{A}^T \underline{P} \underline{l} - \underline{A}^T \underline{P}^T \underline{l} = 0. \quad (7.8)$$

Since $\underline{P}^T = \underline{P}$, we get

$$2(\underline{A}^T \underline{P} \underline{A} \underline{x} - \underline{A}^T \underline{P} \underline{l}) = 0. \quad (7.9)$$

Hence

$$\underline{A}^T \underline{P} \underline{A} \underline{x} = \underline{A}^T \underline{P} \underline{l}. \quad (7.10)$$

The solution is then simply

$$\underline{x} = (\underline{A}^T \underline{P} \underline{A})^{-1} \underline{A}^T \underline{P} \underline{l}. \quad (7.11)$$

In a more general form

$$\underline{x} = \underline{G}^{-1} \underline{g}, \quad (7.12)$$

where $\underline{G} = \underline{A}^T \underline{P} \underline{A}$ and $\underline{g} = \underline{A}^T \underline{P} \underline{l}$.

The above method gives us a practical solution to noisy measurements. The solution is optimized according to the influence of noise.

7.2.2 Sequential adjustment

The goal of this method is to improve the solution of Eq. (7.4) by conducting sequential measurements. Each measurement leads to a new estimation of \underline{x} which will be more and more accurate as we continue. To illustrate this, let us divide the observation model into two sets

$$\begin{cases} \underline{l}_k = \underline{A} \underline{x}_k + \underline{n}_k \\ \underline{l}_{k+1} = \underline{A} \underline{x}_{k+1} + \underline{n}_{k+1} \end{cases}. \quad (7.13)$$

Focusing on the first set, a preliminary solution is given by the Eqs (7.11) and (7.12) as

$$\underline{x}_k = (\underline{A}^T \underline{P} \underline{A})^{-1} \underline{A}^T \underline{P} \underline{l}_k = \underline{G}^{-1} \underline{g}_k. \quad (7.14)$$

The solution \underline{x}_k is computed to limit to influence of the noise \underline{n}_k , ($\underline{l}_k \approx \underline{A} \underline{x}_k$). Next we consider the second set of measurements, and obtain similarly

$$\underline{x}_{k+1} = (\underline{A}^T \underline{P} \underline{A})^{-1} \underline{A}^T \underline{P} \underline{l}_{k+1} = \underline{G}^{-1} \underline{g}_{k+1}. \quad (7.15)$$

Let us assume that the second measurement is a little bit different to the first one

$$\underline{l}_{k+1} = \underline{l}_k + \Delta \underline{l}. \quad (7.16)$$

Eq. (7.15) becomes

$$\underline{x}_{k+1} = \underline{G}^{-1} \underline{A}^T \underline{P} (\underline{l}_k + \Delta \underline{l}). \quad (7.17)$$

Furthermore, using Eq. (7.14)

$$\underline{x}_{k+1} = \underline{x}_k + \underline{G}^{-1} \underline{A}^T \underline{P} \Delta \underline{l}. \quad (7.18)$$

Since $\Delta \underline{l} = \underline{l}_{k+1} - \underline{l}_k$,

$$\underline{x}_{k+1} = \underline{x}_k + \underline{G}^{-1} \underline{A}^T \underline{P} (\underline{l}_{k+1} - \underline{l}_k). \quad (7.19)$$

Finally $\underline{l}_k \approx \underline{A} \underline{x}_k$, and the updated state equation takes the form

$$\underline{x}_{k+1} = \underline{x}_k + \underline{K} (\underline{l}_{k+1} - \underline{A} \underline{x}_k), \quad (7.20)$$

where $\underline{K} = (\underline{A}^T \underline{P} \underline{A})^{-1} \underline{A}^T \underline{P}$.

The above equation is a linear combination of \underline{x}_k . Consequently, $\underline{A} \underline{x}_k$ may be considered as a prediction of \underline{l}_{k+1} before any measurements are conducted. This result is used in Kalman filtering.

7.3 Kalman filtering

7.3.1 Basic concept

The Kalman filter may be explained by the following analogy. A man sleeping in a dark room wakes up in the middle of the night and wants to go out of the room. The room lights are out, so he has to find the exit door in total darkness. He has *a priori* information on where the door is located. He makes an *estimation* based on this information and starts to walk in one direction. After a short while, he walks straight onto a wall. This is regarded as a *measurement*. He now knows a little bit more where he is located in the room. In other words, he *updates* his initial estimate with the measurement. At this point a *prediction* of a new direction is possible. The process is repeated with measurements and will consequently improve his chances of finding the door.

The above is the main essence of Kalman filtering. It is based on sequential adjustment in the static case. All observations up to epoch t are used to obtain optimal estimations of the unknowns. Fortunately, we do not need to store any data as we proceed to the next estimate.

7.3.2 The system model

In a dynamic system such as an airplane, the unknown parameters are the attitude (roll, pitch and yaw), the coordinates and the velocity. These form the elements of the state vector. The state vector is time dependent, and it may be predicted for any instant k by means of system equations. The predicted values are then updated by the use observations, which contain information of the true state vector.

In a discrete form the state vector is defined as $\underline{x}_k^T = [x_1 \cdots x_n]_k$. It consists of n unknown variables. The model for prediction is assumed to be a linear as a function of the previous state

$$\underline{x}_{k+1} = \underline{T}_k \underline{x}_k + \underline{w}_k \quad . \quad (7.21)$$

$\underline{w}_k^T = [w_1 \cdots w_n]_k$ is the noise of the system. The noise of the state vector could for example in an airplane be motor vibrations. Eq. (7.21) is often called the Markov process where \underline{T}_k is the transition matrix.

On the observational side, the *measurement* model is given by Eq. (7.4) as

$$\underline{l}_k = \underline{A}_k \underline{x}_k + \underline{n}_k \quad . \quad (7.22)$$

The measured noise \underline{n}_k is due to the sensors themselves (sensor noise).

7.3.3 The Kalman equations

The updated estimated state vector $\hat{\underline{x}}_k$ is a linear combination with the *a priori* state $\hat{\underline{x}}_k^-$ and the error of measured prediction. The relation was established previously with Eq. (7.20)

$$\hat{\underline{x}}_k = \hat{\underline{x}}_k^- + \underline{K}_k (\underline{l}_k - \underline{A}_k \hat{\underline{x}}_k^-) \quad . \quad (7.23)$$

The main problem in Eq. (7.23) is to find the optimal \underline{K}_k . This is done by minimizing the trace of the error covariance matrix, \underline{Q}_k with respect to \underline{K}_k . This can be done since the trace of \underline{Q}_k is a sum of the mean square errors in the estimates of all the elements of \underline{x}_k

$$\frac{\partial \text{trace}(\underline{Q}_k)}{\partial \underline{K}_k} = 0. \quad (7.24)$$

By definition the error covariance matrix of the updated state vector is

$$\underline{Q}_k \equiv E[(\underline{x}_k - \hat{\underline{x}}_k)(\underline{x}_k - \hat{\underline{x}}_k)^T]. \quad (7.25)$$

Correspondingly, the error covariance matrix of the estimated state vector is

$$\underline{Q}_k^- \equiv E[(\underline{x}_k - \hat{\underline{x}}_k^-)(\underline{x}_k - \hat{\underline{x}}_k^-)^T]. \quad (7.26)$$

\underline{Q}_k can be obtained using Eq. (7.23)

$$\underline{x}_k - \hat{\underline{x}}_k = \underline{x}_k - \hat{\underline{x}}_k^- - \underline{K}_k (\underline{A}_k \underline{x}_k + \underline{n}_k - \underline{A}_k \hat{\underline{x}}_k^-) \quad (7.27)$$

or simply

$$\underline{x}_k - \hat{\underline{x}}_k = (\underline{x}_k - \hat{\underline{x}}_k^-) - \underline{K}_k \underline{A}_k (\underline{x}_k - \hat{\underline{x}}_k^-) - \underline{K}_k \underline{n}_k \quad (7.28)$$

Eqs. (7.27) and (7.28) are inserted into Eq. (7.25) and yields

$$\underline{Q}_k = \underline{Q}_k^- - \underline{Q}_k^- \underline{A}_k^T \underline{K}_k^T - \underline{K}_k \underline{A}_k \underline{Q}_k^- + \underline{K}_k \underline{A}_k \underline{Q}_k^- + \underline{K}_k \underline{A}_k \underline{Q}_k^- \underline{A}_k^T \underline{K}_k^T + \underline{K}_k \underline{R}_k \underline{K}_k^T. \quad (7.29)$$

Or simply

$$\underline{Q}_k = (\underline{Q}_k^- - \underline{K}_k \underline{A}_k \underline{Q}_k^-) (\underline{I} - \underline{A}_k^T \underline{K}_k^T) + \underline{K}_k \underline{R}_k \underline{K}_k^T. \quad (7.30)$$

From the above expression is possible to compute Eq. (7.24). The derivation rules are listed in Appendix E.

$$\frac{\partial [\text{trace}(\underline{Q}_k)]}{\partial \underline{K}_k} = -2 \underline{Q}_k^{-T} \underline{A}_k^T + 2 \underline{K}_k \underline{A}_k \underline{Q}_k^- \underline{A}_k^T + 2 \underline{K}_k \underline{R}_k = 0 \quad (7.31)$$

The optimal \underline{K}_k used in the Eq. (7.23) is then

$$\underline{K}_k = \underline{Q}_k^- \underline{A}_k^T (\underline{A}_k \underline{Q}_k^- \underline{A}_k^T + \underline{R}_k)^{-1} \quad (7.32)$$

This is the same gain factor that we found in the least square adjustment (Eq. 7.20).

Furthermore, the updated error covariance matrix is calculated by inserting Eq. (7.32) in (7.29)

$$\underline{Q}_k = (\underline{I} - \underline{K}_k \underline{A}_k) \underline{Q}_k^- \quad (7.33)$$

The error covariance matrix has decreased. At this point we are finally able to calculate the updated state vector with measurements by Eq. (7.23). The predicted state vector is computed by the model equation

$$\hat{\underline{x}}_{k+1}^- = \underline{T}_k \hat{\underline{x}}_k \quad (7.34)$$

based on the Markov process (7.21), ignoring the model noise. The predicted error covariance matrix before the measurement \underline{l}_{k+1} then becomes

$$\underline{Q}_{k+1}^- = E[(\underline{x}_{k+1} - \hat{\underline{x}}_{k+1}^-)(\underline{x}_{k+1} - \hat{\underline{x}}_{k+1}^-)^T] = \underline{T}_k \underline{Q}_k \underline{T}_k^T + \underline{W}_k \quad (7.35)$$

7.3.4 Kalman algorithm – filter sequence

The filter sequence is based on 4 steps: Estimation – Measurements – Updates and Prediction.

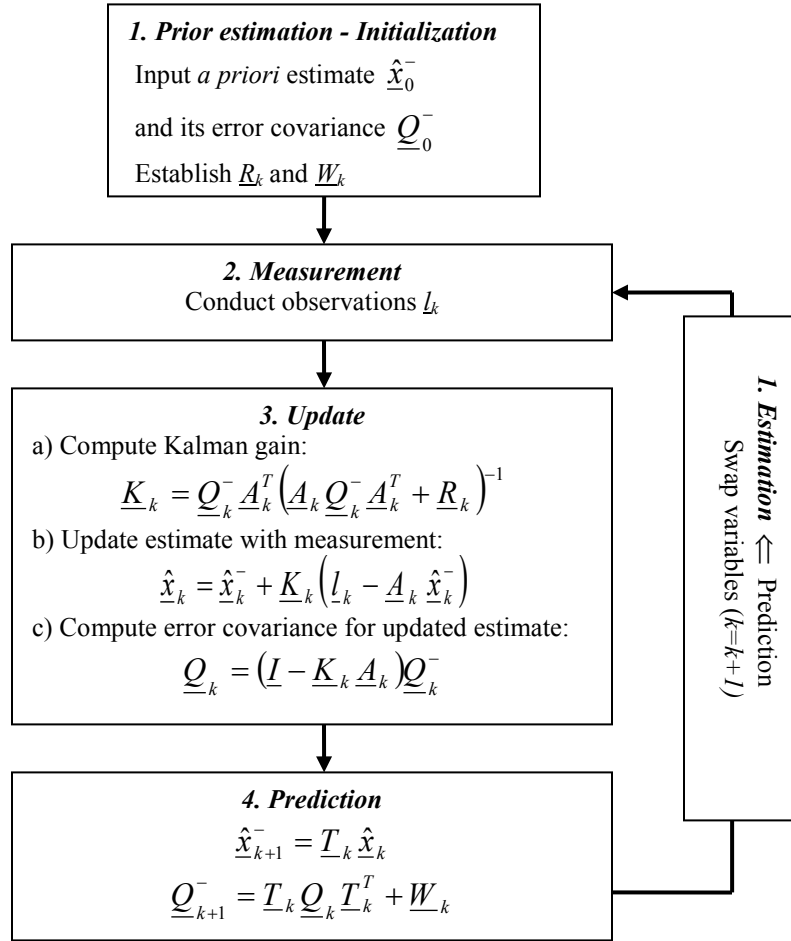


Figure 7.1. The Kalman filter float diagram

7.4 Airborne attitude estimation

7.4.1 Axis configuration

The aim of this study is to obtain an estimate of the attitude to our airborne instruments. The measurements need to include both the acceleration and the turn rate in 3 dimensions (3D).

Fig.7.2 shows the fixed frame coordinate system of the aircraft. The sensors are 3 orthogonal mounted rotational rate sensors (gyros) and 3 accelerometers. The x-axis is out of the nose of the aircraft, the y-axis points out the right wing, and z-axis is down through the center of mass. The roll angle ϕ is defined positive (+) with right wing down.

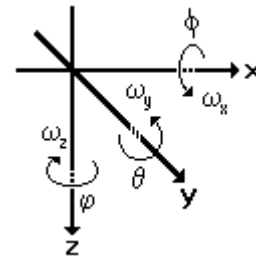


Figure 7.2. Axis. ϕ is roll, θ pitch and φ yaw. ω is the angular turn rate [deg./s].

The pitch angle θ is defined similarly with nose up as positive direction. ϕ , yaw or heading, is positive in the clockwise direction when viewed from above. The angular body rates are as follows. ω_x is the roll rate (angular velocity) about the x-axis with positive direction as right wing down. ω_y is the pitch rate about the y-axis with positive direction defined as nose up. Correspondingly, ω_z is the yaw rate about the z-axis, positive in the clockwise direction when viewed from above.

7.4.2 Measurements

The measured attitude or the pitch, roll and yaw angles of the aircraft are calculated from the accelerometer

$$\underline{l}_k = \begin{bmatrix} \phi \\ \theta \\ \varphi \end{bmatrix}_k = \begin{bmatrix} \text{ArcTan}(-a_y / a_z) \\ \text{ArcSin}(-a_x / \sqrt{a_x^2 + a_y^2 + a_z^2}) \\ \text{Compass} \end{bmatrix}_k, \quad (7.36)$$

where $a_i \forall i \in \{x, y, z\}$ represents the 3 axis accelerometer readings. These angles are also known as the Euler angles. The attitude calculated by Eq. (7.36) assumes a non-accelerated reference frame.

7.4.3 The state vector

In order to obtain smooth and continuous attitude estimates as a function of time, quaternion mathematics is used. Quaternions make rotations in four dimensions. 3D rotations by Euler angles may suffer from an effect called ‘‘Gimbal lock’’. This problem occurs when the final rotational matrix is created by multiplying the rotational matrices in the x, y and z-axis, respectively. The order of multiplication becomes very important. The problem is fixed by the use of a quaternion as the state vector

$$\underline{q}_k^T = [q_s \ q_x \ q_y \ q_z]_k. \quad (7.37)$$

A short summary of the nature and behavior of quaternions are found in Appendix D.

7.4.4 The measurement model

The measurements are related to the observational model as follows

$$\underline{l}_k = \begin{bmatrix} \phi \\ \theta \\ \varphi \end{bmatrix}_k = \underline{A}_k \cdot \begin{bmatrix} q_s \\ q_x \\ q_y \\ q_z \end{bmatrix}_k + \underline{n}_k = \underline{A}_k \cdot \underline{q}_k + \underline{n}_k. \quad (7.38)$$

The measurement matrix \underline{A}_k is defined as

$$\underline{A}_k \cong \begin{bmatrix} \frac{\partial \phi}{\partial q_s} & \frac{\partial \phi}{\partial q_x} & \frac{\partial \phi}{\partial q_y} & \frac{\partial \phi}{\partial q_z} \\ \frac{\partial \theta}{\partial q_s} & \frac{\partial \theta}{\partial q_x} & \frac{\partial \theta}{\partial q_y} & \frac{\partial \theta}{\partial q_z} \\ \frac{\partial \varphi}{\partial q_s} & \frac{\partial \varphi}{\partial q_x} & \frac{\partial \varphi}{\partial q_y} & \frac{\partial \varphi}{\partial q_z} \end{bmatrix}_k \quad (7.39)$$

This matrix is also called a transition matrix since it permits us to convert from quaternion- to Euler frames of reference. The mathematics behind the construction of \underline{A}_k is non- trivial, but doable. It is a Jacobian matrix.

7.4.5 The system model - prediction

The instantaneous rate change of an airborne quaternion is given by Hughes ²⁹.

$$\dot{\underline{q}}_k = \frac{1}{2} \underline{\Omega}_k \cdot \underline{q}_k, \quad (7.40)$$

where $\underline{\Omega}_k$ is the quaternion omega matrix for the current rotational sensors

$$\underline{\Omega}_k = \begin{bmatrix} 0 & -\omega_x & -\omega_y & -\omega_z \\ \omega_x & 0 & \omega_z & -\omega_y \\ \omega_y & -\omega_z & 0 & \omega_x \\ \omega_z & \omega_y & -\omega_x & 0 \end{bmatrix}_k. \quad (7.41)$$

$(\omega_x, \omega_y, \omega_z)_k$ represent the rolling, pitching and yawing turn rates in the strap down inertial reference frame, as measured by the rotational rate sensors at the moment \underline{t}_k .

The transition matrix \underline{T}_k is then given as

$$\underline{T}_k = \underline{I} + \frac{1}{2} \underline{\Omega}_k \cdot \Delta t_k, \quad (7.42)$$

where $\Delta t_k = t_{k+1} - t_k$ is the epoch or sampling period. The transition matrix satisfies the relation

$$\underline{q}_{k+1} = \underline{T}_k \underline{q}_k + \underline{w}_k, \quad (7.43)$$

which is the Markov process or the predictive state of the Kalman filter. Note that the transition matrix is updated at every epoch by the use of measurements from the gyros.

The fundamental Kalman relations are now established by Eqs. (7.38) and (7.43). The final filter sequence is shown in Fig. 7.3. The transition matrix Eq. (7.42) is based on a semi-continuous predictive approach - described in Appendix F.

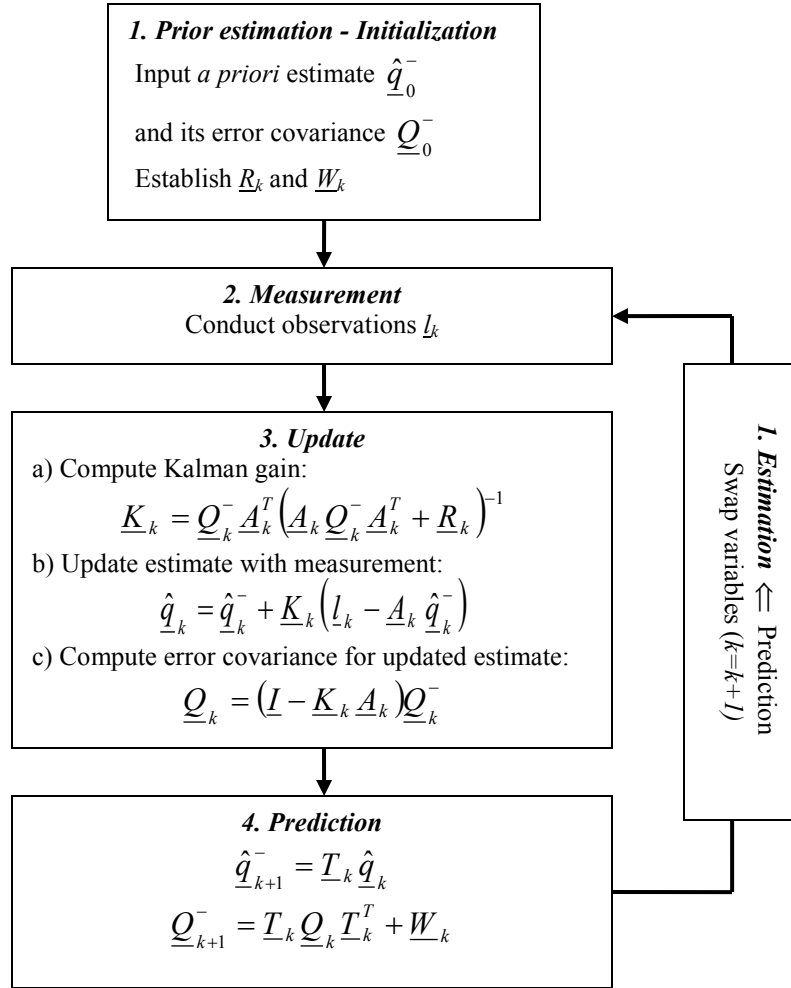


Figure 7.3. The Kalman filter float diagram for airborne attitude control.

\underline{W}_k	Noise model covariance matrix
\underline{R}_k	Noise measurement covariance matrix
\underline{l}_k	Measurements (Euler angle)
\underline{K}_k	Kalman gain
$\hat{\underline{q}}_k$	Updated estimation with measurements \underline{l}_k
\underline{Q}_k	Updated error covariance matrix
$\hat{\underline{q}}_k^-$	Estimation before measurement \underline{l}_k
\underline{Q}_k^-	Estimation error covariance matrix before measurement \underline{l}_k
$\hat{\underline{q}}_{k+1}^-$	Prediction before measurement \underline{l}_{k+1}
\underline{Q}_{k+1}^-	Prediction error covariance matrix before measurement \underline{l}_{k+1}
\underline{T}_k	Markov matrix
\underline{A}_k	Measurement matrix

7.5. Attitude applications

7.5.1 Experimental setup

The first test of the Kalman filter was conducted during the AIRSPEX MMVI campaign carried out as a part of the course AGF331 - Remote sensing and spectroscopy, at the University Centre in Svalbard (UNIS). The platform contains an IMU sensor including turn rate sensors, accelerometers and magnetometers acquired from the American firm Microstrain Inc. The GPS is from the company Garmin. Fig. 7.4 shows the setup.

The hyperspectral imager and the attitude sensors were mounted on Dornier 228-202K airplane operated by the commercial company Lufttransport AS. The instruments were looking out of a sliding door at the side of the airplane, pointing 30 degrees off nadir.

The flight was carried out at an altitude of 1500 m above sea level 3rd of May 2006, traversing Longyearbyen (78°N, 15°E), Norway, from south-west to north-east.

Fig. 7. 5 shows the results of image corrections with the use of Kalman filtered data from the IMU / GPS.



Figure 7.4. The GPS is linked to the laptop via a USB hub. The turn rate sensors / accelerometers / magnetometers are connected to the laptop with the same hub. The sensors are powered by +12VDC battery.

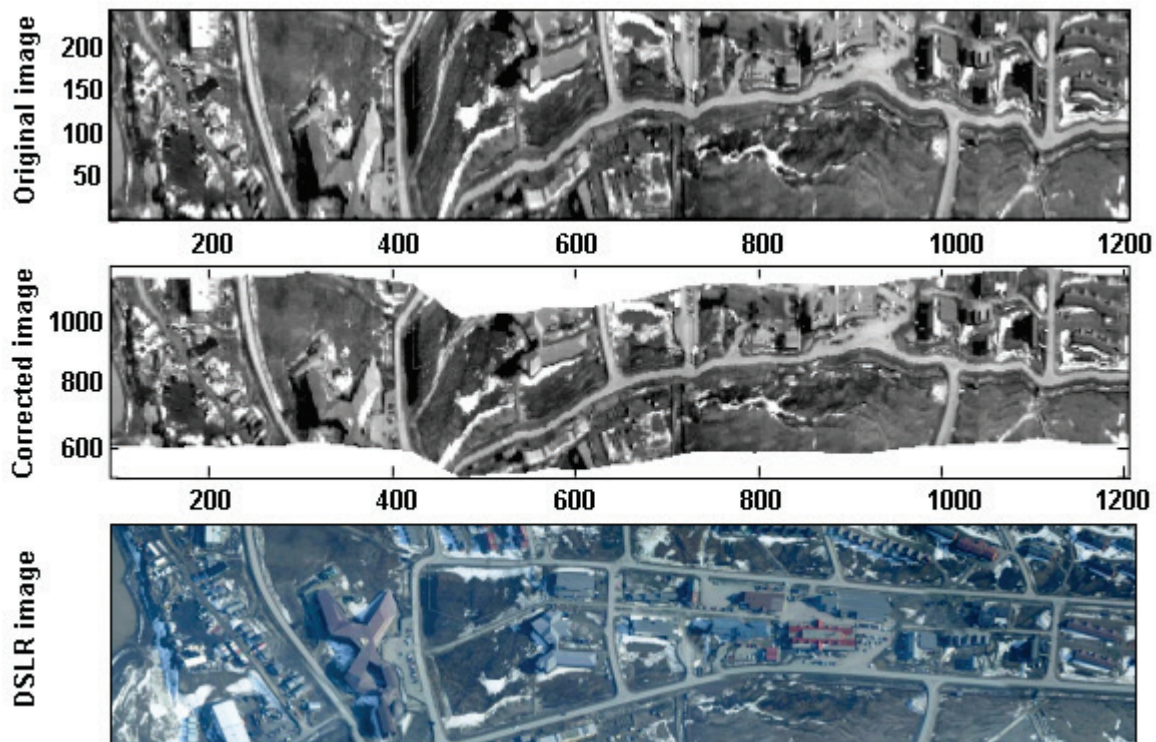


Figure 7.5. Kalman attitude corrected hyperspectral image over Longyearbyen, Norway, 03.05.2006.

Top image is raw data from the hyperspectral imager at wavelength 550 nm. Middle image is the attitude corrected one. Bottom image is a DSLR image taken prior to the spectral sampling.

It is clear comparing the DSLR image with the corrected one that more work need to be done to make the filter work better.

Reference

1. Alonso / Finn, Fundamental University Physics, Second Edition, Volume II, Fields and Waves, Addison - Wesley publishing company, ISBN 0-201-00162-4, 1967.
2. E. Hecht, Theory and Problems of Optics, Schaum's outline series, McGraw-Hill, Inc., ISBN 0-07-027730-3, 1975.
3. C. Palmer, Diffraction Gratings Handbook, 5th ed. Thermo RGL (Richardson Grating Laboratory, 2002), pp. 142 – 144.
4. J. M. Lerner and A. Thevenon, The optics of spectroscopy, A tutorial v2.0, Jobin Yvon Instruments SA, Inc., 1988.
5. H. Ebert, Zwei Formen von Spectrographen, Annalen der Physik., 38, 489-493, ISSN 0003-3804, 1889.
6. W. G. Fastie, A Small Plane Grating Monochromator, J. Opt. Soc. Am., 42, 641, 1952.
7. W. G. Fastie, Image Forming Properties of the Ebert Monochromator, J. Opt. Soc. Am., 42, 647, 1952.
8. M. Czerny and A. F. Turner, Über den Astigmatismus bei Spiegelspektrometern, Zeitschrift fuer Physik, 792-797, Berlin, ISSN 0044-3328, 1930.
9. M. Aikio, Hyperspectral prism-grating-prism imaging spectrograph, VTT Publications, 435, 15-114, 2001.
10. K. Henriksen, Nordlysspektrofotometeret SP3, Institutt for matematisk realfag, Universitetet i Tromsø, Mai 1977.
11. G. Vane, ed., Imaging Spectroscopy II, Proc. SPIE 834, 1988.
12. W.L. Wolfe, Introduction to Imaging Spectrometers, Vol. TT25 of Tutorial Text Series, SPIE Press, Bellingham, Wash., pp. 1-147, 1997.
13. E. Herrala, and J. Okkonen, Imaging spectrograph and camera solutions for industrial applications, International Journal of Pattern Recognition and Artificial Intelligence, 10 (1), 43-54, 1996.
14. T. Hyvarinen, E. Herrala, and A. Dall'Ava, Direct sight imaging spectrograph: a unique add-on component brings spectral imaging to industrial applications, SPIE Symposium on Electronic Imaging, Paper 3302-21, San Jose, California, January 25-30, 1998.
15. F. Sigernes, K. Heia, H. Nilsen, and T. Svenøe, Imaging spectroscopy applied in the fish industry?, Norwegian Society for Image Processing and Pattern Recognition, 2, 16-24, 1998.

16. Nilsen, H., Esaiassen, M., Karsten, H., and Sigernes, F., VIS/NIR spectroscopy - a new tool for the evaluation of fish freshness?, *Journal of Food Science*, 67, No. 5, 1821 - 1826, 2002.
17. Sigernes, F., Lorentzen, D.A., Heia, K., and Svenøe, T., A multi-purpose spectral imager, 39, NO.18, 3143, (front cover issue), *Applied Optics*, 2000.
18. W. Niblack, *An Introduction to digital image processing*, Vol. 1, Prentice / Hall international (UK) Ltd., pp. 168-181, 1968.
19. Lambert, J.H., "Photometria, sive de mensura et gradibus luminis colorum et umbrae", Augsburg, 1760. German trans. by E. Anding in Ostwald's "Klassiker d. Exacten Wissenschaften" (Engelmann, Leipzig, 1892) p. 324.
20. Oriel Instruments, *Instruction Manual, Quartz tungsten halogen lamp standards of spectral irradiance: 1kW, 200W, 45W, 63350-M*, p. 6, 250 Long Beach Boulevard, Stratford, CT 06497-0872, USA, 1996.
21. Chamberlain, J. W., *Physics of the Aurora and Airglow*, 569 – 573, Academic Press, New York and London, 1961.
22. Baker, D. and G. Romick, The rayleigh: interpretation of the unit in terms of column emission rate or apparent radiance expressed in SI units, *Applied Opt.*, 15(8): p. 1966-1968, 1976.
23. Walker, J.H., Saunders, R.D., Jackson, J.K., and Mielenz, K.D., Results of a CCPR intercomparison of spectral irradiance measurements by National Laboratories, *J. Res. Natl. Inst. Standard Technol.*, 96, 647-668, 1991.
24. Gröbner, J., Rembges, D., Bais, A.F., Blumthaler, M., Cabot, T., Josefsson, W., Koskela, T., Thorseth, T.M., Webb, A.R., and Wester, U., Quality assurance of reference standards from nine European solar-ultraviolet monitoring laboratories, *Applied Optics*, 41, 21, 4278 – 4282, 2002.
25. Gröbner, J. and P. Sperfeld, Direct traceability of the portable QASUM irradiance scale to the primary irradiance standard of the PTB, *Metrologia*, 42, 134-139, 2005.
26. Hofmann-Wellenhof, B., and Lichtenegger, H., *GPS Theory and Practice*, 5th ed. Springer-Verlag Wien New York, ISBN 3-211-83534-2, p.248 – 254, 1992.
27. Brown, R.G. and Hwang, P.Y.C., *Introduction to random signals and applied Kalman filtering*, third edition, New York: John Wiley & Sons, ISBN 0-471-12839-2, p.214 – 219, 1997.
28. Teunissen, PJG, The invertible GPS ambiguity transformations, *Manuscripta geodaetica*, 20, 489-497, 1995.

29. Hughes, P.C., *Spacecraft Attitude Dynamics*, New York: John Wiley & Sons, 1986.
30. Hamilton, W. R., *Elements of Quaternions*, Longmans, Green & Co., London, 1866.
31. Klumpp, A.R.: Singularity-Free Extraction of a Quaternion from a Direction-Cosine Matrix, *Journal of Spacecraft*, Vol. 13, No 12, pp. 754-755, Dec. 1976.
32. Sigernes, F., Holmes, J. M., Dyrland, M., Lorentzen, D.A., Chernous, S., Svenøe, S., Moen, J., and Deehr, C. S., Absolute calibration of optical devices with a small field of view, *Journal of Optical Technology* **74**, 669-674, 2007.
33. Sigernes, F., Holmes, J. M., Dyrland, M., Lorentzen, D. A., Svenøe, T., Heia, K., Aso, T., Chernouss, S., and Deehr, C. S., Sensitivity calibration of digital colour cameras for auroral imaging, *Optics Express*, Vol. 16, Issue 20, pp. 15623-15632, 2008.
34. Sigernes, F., Dyrland, M., Peters, N., Lorentzen, D. A., Svenøe, T., Heia, K., Chernouss, S., Deehr, C. S., and Kosch, M., The absolute sensitivity of digital colour cameras, *Optics Express*, Vol. 17, Issue 22, pp. 20211-20220, 2009.
35. Berry, R., and Burnell, J., Measuring CCD performance in *Handbook of Astronomical Image Processing*, (Willmann-Bell, Inc., 2006), pp. 227-248.

APPENDIX A

Basic Ebert-Fastie equations

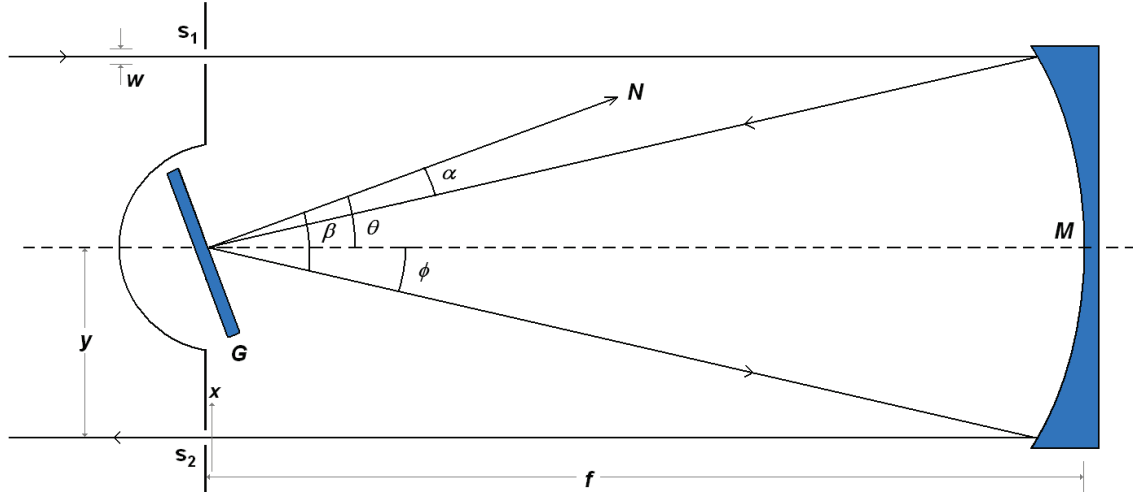


Figure A.1. Optical diagram Ebert- Fastie spectrometer. G is plane reflective grating, S₁ entrance slit, S₂ exit slit, and M concave mirror. N is grating normal, θ grating angle, ϕ instrumental constant ($\tan \phi = y/f$), f focal length, α incident angle, β diffracted angle, y $\frac{1}{2}$ slit – slit distance, a groove spacing, n spectral order, and w slit width.

The grating equation is

$$n\lambda = a (\sin \alpha + \sin \beta), \text{ where } \alpha = \theta - \phi \text{ and } \beta = \theta + \phi.$$

Then $n\lambda = a [(\sin \theta \cos \phi - \sin \phi \cos \theta) + (\sin \theta \cos \phi + \sin \phi \cos \theta)]$ or

$$\Rightarrow n\lambda = 2a \sin \theta \cos \phi.$$

Angular dispersion is

$$\frac{d}{d\beta}(n\lambda) = a \cos \beta,$$

and since $dx = f d\beta$ then linear dispersion becomes

$$\frac{d\lambda}{dx} = \frac{d\lambda}{d\beta f} = \frac{a \cos \beta}{nf} = \frac{a \cos(\theta + \phi)}{nf}$$

The theoretical bandpass of the instrument is then defined as

$$BP = FWHM = \frac{d\lambda}{dx} \times w = \frac{a \cos(\theta + \phi)}{nf} \times w.$$

APPENDIX B

Basic calculations for a 400 – 900 nm GRISM spectrograph

(1) Borate flint glass

Wavelength λ [nm]	Refractive index n	Diffracted angle β [deg.]	Linear dispersion $d\lambda/dx$ [nm/mm]
400	1.58942	25.0630	99.6853
500	1.57606	22.8950	110.823
600	1.56880	20.8891	117.220
700	1.56442	18.9694	121.303
800	1.56158	17.1031	124.153
900	1.55963	15.2734	126.278

(2) BK7

Wavelength λ [nm]	Refractive index n	Diffracted angle β [deg.]	Linear dispersion $d\lambda/dx$ [nm/mm]
400	1.53144	23.8148	105.426
500	1.52152	21.7397	114.634
600	1.51613	19.7884	119.882
700	1.51288	17.9048	123.264
800	1.51078	16.0643	125.660
900	1.50933	14.2540	127.471

Table B.1. Linear dispersion using a grism with $\omega = \alpha = 20^\circ$, grating groove spacing $a = 3333.33$ nm (a 300 lines / mm), spectral order $n = 1$, and a detector lens with focal length of $f_3 = 25$ mm. Cauchy's index of refraction constants are $A_1 = 1.5523$ and $B_1 = 5939.39$ nm for Borate flint glass in panel (1), and $A_1 = 1.50389$ and $B_1 = 4407.54$ nm in panel (2) for BK7.

Wavelength λ [nm]	Slit width magnification	Bandpass BP [nm]
400	0.617	1.54
500	0.607	1.68
600	0.599	1.75
700	0.591	1.79
800	0.585	1.82
900	0.580	1.83

Table B.2. Slit width magnification and theoretical spectral bandpass of a grism spectrograph. The prism is of Borate flint glass with $\omega = \alpha = 20^\circ$. The transmitting grating has 300 lines / mm. The spectral order is $n = 1$. The lens between the entrance slit and the grism has $f_2 = 42$ mm as focal length. The detector lens has a focal length of $f_3 = 25$ mm. The entrance slit width is $w = 0.025$ mm.

APPENDIX C

Spatial resolution calculations with different front optics

z[m]	dx [m]	Δx [m]	Δy [m]	SW [m]
3000	6.25	6.71	8.65	1124.5
2500	5.21	5.67	7.21	937.3
2000	4.17	4.63	5.77	750.1
1500	3.13	3.59	4.33	562.9
1000	2.08	2.54	2.88	374.4
850	1.77	2.23	2.45	318.5
500	1.04	1.50	1.44	187.2
400	0.83	1.07	1.15	150

Table C.1.

$f_1=12$ mm

$v = 100$ km/hr (27.78 m/s)

$w = 0.025$ mm

$\frac{1}{2}$ " CCD (320 x 240 pixels)

25 frames per second.

$N=130$ and $\Delta t = 8.33$ ms

$\tau = 31.667$ ms

Readout distance: 0.88 m

Critical altitude ~ 400 m

z[m]	dx [m]	Δx [m]	Δy [m]	SW [m]
3000	4.69	4.92	6.49	844
2500	3.91	4.14	5.40	703
2000	3.13	3.36	4.33	563
1500	2.34	2.58	3.25	422
1000	1.56	1.79	2.16	281
850	1.39	1.56	1.84	239
600	0.94	1.17	1.30	169
500	0.78	1.01	1.08	140

Table C.2.

$f_1=16$ mm

$v = 100$ km/hr (27.78 m/s)

$w = 0.025$ mm

$\frac{1}{2}$ " CCD (320 x 240 pixels)

25 frames per second.

$N=130$ and $\Delta t = 8.33$ ms

$\tau = 31.667$ ms

Readout distance: 0.88 m

Critical altitude $\sim 500 - 600$ m

z[m]	dx [m]	Δx [m]	Δy [m]	SW [m]
3000	3.0	3.23	4.15	540
2500	2.5	2.73	3.46	450
2000	2.0	2.23	2.77	360
1500	1.5	1.73	2.08	270
1000	1.0	1.23	1.38	180
850	0.85	1.08	1.18	153
500	0.5	0.73	0.69	90

Table C.3.

$f_1=25$ mm

$v = 100$ km/hr (27.78 m/s)

$w = 0.025$ mm

$\frac{1}{2}$ " CCD (320 x 240 pixels)

25 frames per second.

$N=130$ and $\Delta t = 8.33$ ms

$\tau = 31.667$ ms

Readout distance: 0.88 m

Critical altitude ~ 850 m

APPENDIX D

Quaternion mathematics

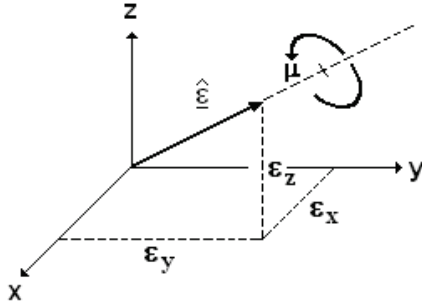


Figure D.1. Euler's theorem

Any fixed point rotation of a rigid body can be described as a single rotation of a certain angle (the eigenangle, μ) about a single axis (the eigenaxis of Euler axis, \hat{e}). This is known as Euler's theorem.

Based on the Euler's theorem, Hamilton³⁰ constructed the quaternion. For our work, only a subset of Hamilton's quaternion calculus is required to do rotations.

In the following we write a quaternion as a four element vector

$$\underline{q} = [q_s, q_x, q_y, q_z], \quad (\text{D.1})$$

where $q_s = \cos(\mu/2)$, $q_x = \epsilon_x \sin(\mu/2)$, $q_y = \epsilon_y \sin(\mu/2)$ and $q_z = \epsilon_z \sin(\mu/2)$.

It can also be viewed as a scalar coupled to a three-dimensional vector

$$\underline{q} = (q_s, [\vec{q}]) = (q_s, \vec{q}). \quad (\text{D.2})$$

The magnitude of the quaternion is then defined as

$$\|\underline{q}\| = \sqrt{q_s^2 + q_x^2 + q_y^2 + q_z^2}. \quad (\text{D.3})$$

The dot product between the quaternions \underline{p} and \underline{q} is

$$\underline{p} \cdot \underline{q} = p_s q_s + p_x q_x + p_y q_y + p_z q_z. \quad (\text{D.4})$$

From (D.2) and (D.3) we see that

$$\|\underline{q}\|^2 = \underline{q} \cdot \underline{q}. \quad (\text{D.5})$$

The quaternion is normalized if the magnitude equals 1. Therefore, normalization of a quaternion is conducted by dividing each component with its magnitude.

\underline{q}_N is the unit quaternion of \underline{q} given by the expression:

$$\underline{q}_N = \frac{\underline{q}}{\|\underline{q}\|} = \frac{[q_s, q_x, q_y, q_z]}{\sqrt{q_s^2 + q_x^2 + q_y^2 + q_z^2}}. \quad (\text{D.6})$$

Furthermore, the conjugate of a quaternion is defined as

$$\bar{\underline{q}} = [q_s, -q_x, -q_y, -q_z]. \quad (\text{D.7})$$

The multiplication of two quaternions under the above rules leads to the definition

$$\underline{p} \circ \underline{q} = [p_s q_s - \vec{p} \cdot \vec{q}, p_s \vec{q} + q_s \vec{p} + \vec{p} \times \vec{q}]. \quad (\text{D.8})$$

A vector \underline{v} can be rotated rather elegant by the use of a quaternion. We assume that \underline{q} is a rotation unit quaternion.

$$\underline{v}' = \underline{q} \circ \underline{v} \circ \underline{q}^{-1} \quad (\text{D.9})$$

The above expression is much faster to implement than the transformation using rotational matrices. Note that the inverse is equivalent to calculate the conjugate of the quaternion.

The eigenaxis and the eigenangle can be computed by the following procedure known as Hall's method

$$\begin{aligned}
 q_m &= (q_x^2 + q_y^2 + q_z^2) \\
 \text{sign}(x) &= \begin{cases} +1 & x > 0 \\ +0 & x = 0 \\ -1 & x < 0 \end{cases} \\
 \mu &= 2 \text{ ArcCos}(q_m) \\
 \underline{\hat{\varepsilon}} &= \frac{\text{sign}(q_s)}{q_m} [q_x, q_y, q_z].
 \end{aligned} \tag{D.10}$$

The above method will fail if $q_s = 0$.

Traditionally, the angles representing the pitch, roll and yaw (ϕ, θ, φ) are the most used parameters to describe rotations of an aircraft. We call them the Euler angles. A set of Euler angles can be converted to a quaternion by the following equations

$$\begin{aligned}
 q_s &= \text{Cos}(\phi/2)\text{Cos}(\theta/2)\text{Cos}(\varphi/2) + \text{Sin}(\phi/2)\text{Sin}(\theta/2)\text{Sin}(\varphi/2) \\
 q_x &= \text{Sin}(\phi/2)\text{Cos}(\theta/2)\text{Cos}(\varphi/2) - \text{Cos}(\phi/2)\text{Sin}(\theta/2)\text{Sin}(\varphi/2) \\
 q_y &= \text{Cos}(\phi/2)\text{Sin}(\theta/2)\text{Cos}(\varphi/2) + \text{Sin}(\phi/2)\text{Cos}(\theta/2)\text{Sin}(\varphi/2) \\
 q_z &= \text{Cos}(\phi/2)\text{Cos}(\theta/2)\text{Sin}(\varphi/2) - \text{Sin}(\phi/2)\text{Sin}(\theta/2)\text{Cos}(\varphi/2).
 \end{aligned} \tag{D.11}$$

For small angles we obtain

$$\begin{aligned}
 q_s &\approx 1 \\
 q_x &\approx \phi/2 = \text{roll}/2 \\
 q_y &\approx \theta/2 = \text{pitch}/2 \\
 q_z &\approx \varphi/2 = \text{yaw}/2
 \end{aligned} \tag{D.12}$$

Similarly, a quaternion is converted back to Euler angles by

$$\begin{aligned}
 \phi &= \text{ArcTan}(2(q_y q_z + q_s q_x)/(1 - 2(q_x^2 + q_y^2))) \\
 \theta &= -\text{ArcSin}(2(q_x q_z - q_s q_y)) \\
 \varphi &= \text{ArcTan}(2(q_x q_y + q_s q_z)/(1 - 2(q_y^2 + q_z^2)))
 \end{aligned} \tag{D.13}$$

A quaternion \underline{q} is called rotational if its normalized and can be converted into a 3×3 rotational matrix DCM. This matrix is often called the Directional Cosine Matrix by Klumpp³¹.

$$\underline{DCM} = \underline{R} = \begin{bmatrix} 1 - 2 \times (q_y^2 + q_z^2) & 2 \times (q_x q_y + q_s q_z) & 2 \times (q_x q_z - q_s q_y) \\ 2 \times (q_x q_y - q_s q_z) & 1 - 2 \times (q_x^2 + q_z^2) & 2 \times (q_y q_z + q_s q_x) \\ 2 \times (q_x q_z + q_s q_y) & 2 \times (q_y q_z - q_s q_x) & 1 - 2 \times (q_x^2 + q_y^2) \end{bmatrix}. \tag{D.14}$$

The inverse procedure, to calculate the quaternion from a rotational matrix, is performed in several stages. If the trace $T = 1 + R_{00} + R_{11} + R_{22}$ of the matrix is greater zero then

$$\begin{aligned}
 S &= 2\sqrt{T} \\
 q_s &= S/4 \\
 q_x &= (R_{12} - R_{21})/S \\
 q_y &= (R_{20} - R_{02})/S \\
 q_z &= (R_{01} - R_{10})/S.
 \end{aligned} \tag{D.15}$$

On the other hand, if T is zero, then we have to identify which major diagonal element that has the greatest value. In the case where $R_{00} > R_{11}$ and $R_{00} > R_{22}$ then

$$\begin{aligned}
 S &= 2\sqrt{(1 + R_{00} - R_{11} - R_{22})} \\
 q_s &= (R_{12} - R_{21})/S \\
 q_x &= S/4 \\
 q_y &= (R_{01} + R_{10})/S \\
 q_z &= (R_{20} + R_{02})/S.
 \end{aligned} \tag{D.16}$$

Else if $R_{11} > R_{22}$ then

$$\begin{aligned}
 S &= 2\sqrt{(1 + R_{11} - R_{00} - R_{22})} \\
 q_s &= (R_{20} - R_{02})/S \\
 q_x &= (R_{01} + R_{10})/S \\
 q_y &= S/4 \\
 q_z &= (R_{12} + R_{21})/S.
 \end{aligned} \tag{D.17}$$

Finally, if the above fails

$$\begin{aligned}
 S &= 2\sqrt{(1 + R_{22} - R_{00} - R_{11})} \\
 q_s &= (R_{01} - R_{10})/S \\
 q_x &= (R_{20} + R_{02})/S \\
 q_y &= (R_{12} + R_{21})/S \\
 q_z &= S/4.
 \end{aligned} \tag{D.18}$$

The translation matrix \underline{A} that translate the quaternion frame to Euler angles is calculated from the estimated state vector \underline{q} and the Directional Cosine Matrix $\underline{R=DCM}$.

$$\underline{A} = \begin{bmatrix} \frac{\partial \phi}{\partial q_s} & \frac{\partial \phi}{\partial q_x} & \frac{\partial \phi}{\partial q_y} & \frac{\partial \phi}{\partial q_z} \\ \frac{\partial \theta}{\partial q_s} & \frac{\partial \theta}{\partial q_x} & \frac{\partial \theta}{\partial q_y} & \frac{\partial \theta}{\partial q_z} \\ \frac{\partial \varphi}{\partial q_s} & \frac{\partial \varphi}{\partial q_x} & \frac{\partial \varphi}{\partial q_y} & \frac{\partial \varphi}{\partial q_z} \end{bmatrix} = \begin{bmatrix} \frac{2R_{22}q_x}{R_{22}^2 + R_{12}^2} & \frac{2(R_{22}q_s + 2R_{12}q_x)}{R_{22}^2 + R_{12}^2} & \frac{2(R_{22}q_z + 2R_{12}q_y)}{R_{22}^2 + R_{12}^2} & \frac{2R_{22}q_y}{R_{22}^2 + R_{12}^2} \\ \frac{2q_y}{\sqrt{1 - R_{02}^2}} & \frac{-2q_z}{\sqrt{1 - R_{02}^2}} & \frac{2q_s}{\sqrt{1 - R_{02}^2}} & \frac{-2q_x}{\sqrt{1 - R_{02}^2}} \\ \frac{2R_{00}q_z}{R_{00}^2 + R_{01}^2} & \frac{2R_{00}q_y}{R_{00}^2 + R_{01}^2} & \frac{2(R_{00}q_x + 2R_{01}q_y)}{R_{00}^2 + R_{01}^2} & \frac{2(R_{00}q_s + 2R_{01}q_z)}{R_{00}^2 + R_{01}^2} \end{bmatrix}$$

(D.19)

APPENDIX E

Matrix derivation

It is forbidden to do

$$\frac{\partial \underline{x}}{\partial \underline{x}} \quad \text{nor} \quad \frac{\partial \underline{x}^T}{\partial \underline{x}^T} . \quad (\text{E.1})$$

On the other hand, it is allowed to

$$\frac{\partial \underline{x}^T}{\partial \underline{x}} = \frac{\partial \underline{x}}{\partial \underline{x}^T} = \underline{I} \quad (\text{E.2})$$

$$\frac{\partial [\underline{P}\underline{x}]}{\partial \underline{x}} = \frac{\partial [\underline{P}\underline{x}]^T}{\partial \underline{x}} = \underline{P}^T \quad (\text{E.3})$$

$$\frac{\partial [\underline{P}\underline{x}]}{\partial \underline{x}^T} = \underline{P} \quad (\text{E.4})$$

In addition if $\underline{C} = \underline{C}^T$ then

$$\frac{\partial [\underline{A}\underline{C}\underline{A}^T]}{\partial \underline{A}} = [\underline{C}\underline{A}^T]^T + \underline{A}\underline{C} = 2 \cdot \underline{A}\underline{C} \quad (\text{E.5})$$

If $\underline{A}\underline{B}$ is a square and \underline{C} is symmetric matrix then

$$\frac{\partial [\text{trace}(\underline{A}\underline{B})]}{\partial \underline{A}} = \underline{B}^T \quad (\text{E.6})$$

$$\frac{\partial [\text{trace}(\underline{A}\underline{C}\underline{A}^T)]}{\partial \underline{A}} = 2\underline{A}\underline{C} \quad (\text{E.7})$$

Derivation of a scalar with a matrix:

$$\frac{\partial \alpha}{\partial \underline{A}} = \left[\frac{\partial \alpha}{\partial a_{ij}} \right] \quad (\text{E.8})$$

APPENDIX F

The Semi-continuous prediction model

In most cases the time dependent state vector $\underline{x}(t)$ of a dynamic system may be modeled by a set of differential equations of the first order

$$\dot{\underline{x}}(t) = \underline{F}(t) \underline{x}(t) + \underline{w}(t) \quad (\text{F.1})$$

$\dot{\underline{x}}(t)$ is the time derivative of the state vector, $\underline{F}(t)$ the dynamic matrix, and $\underline{w}(t)$ is the driving noise. If we assume that the state vector $\underline{x}(t_0)$ is known at initial epoch t_0 and that the dynamic matrix $\underline{F}(t)$ contains only constant coefficients, then the solution for the system equation (F.1) can be written as

$$\begin{aligned} \underline{x}(t) &= \underline{T}(t, t_0) \underline{x}(t_0) + \int_{t_0}^t \underline{T}(t, \tau) \underline{w}(\tau) d\tau \\ &= \underline{T}(t, t_0) \underline{x}(t_0) + \underline{e}(t) . \end{aligned} \quad (\text{F.2})$$

The matrix \underline{T} is called the transition matrix. This matrix can be expressed as a function of the dynamics matrix \underline{F} through a Taylor expansion of $\underline{x}(t)$

$$\underline{x}(t) = \underline{x}(t_0) + \dot{\underline{x}}(t_0)(t - t_0) + \frac{1}{2} \ddot{\underline{x}}(t_0)(t - t_0)^2 + \dots \quad (\text{F.3})$$

Using equation (F.1) and neglecting the driving noise $\underline{w}(t)$

$$\underline{x}(t) = \underline{x}(t_0) + \underline{F}(t_0) \underline{x}(t_0) (t - t_0) + \frac{1}{2} \underline{F}(t_0)^2 \underline{x}(t_0) (t - t_0)^2 + \dots \quad (\text{F.4})$$

In addition, if we introduce $\Delta t = t - t_0$, the transition matrix is given as an infinite series with respect to \underline{F} as

$$\begin{aligned} \underline{T}(t, t_0) &= \underline{I} + \underline{F}(t_0) \Delta t + \frac{1}{2} \underline{F}(t_0)^2 \Delta t^2 + \dots \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \underline{F}(t_0)^n \Delta t^n . \end{aligned} \quad (\text{F.5})$$

In a discrete form Eq. (F.2) may be written as

$$\underline{x}_{k+1} = \underline{T}_k \underline{x}_k + \underline{w}_k , \quad (\text{F.6})$$

where $t_k = k \cdot \Delta t$ and from Eq. (D.5)

$$\underline{T}_k = \underline{I} + \underline{F}_k \cdot \Delta t \quad (\text{F.7})$$

for $n=1$. Eq. (F.6) is known as the Markov process and it is the fundamental system model of prediction in the discrete Kalman filter.