Computer Vision 1 - Lab 1

Pauliuc, Andrei-Sebastian
andrei.pauliuc@student.uva.nl
11596619

van Slooten, Renzo
renzovanslooten@msn.com
11129778

1 Photometric Stereo

1.1 Question - 1

In albedo image we expect to see the Figure 5.11 from chapter 5, which is the reflectance of the surface. This image can be seen as the true colour of the object being photographed. Our results, shown in figure 1, show the object as being more luminous than the original one, the white colour showing a much more intense brightness in the centre of the figure. The cause of this outcome could be the difference between the given input images and the ones used in the book, especially for picture *Sphere_0.0_0.0.png*.

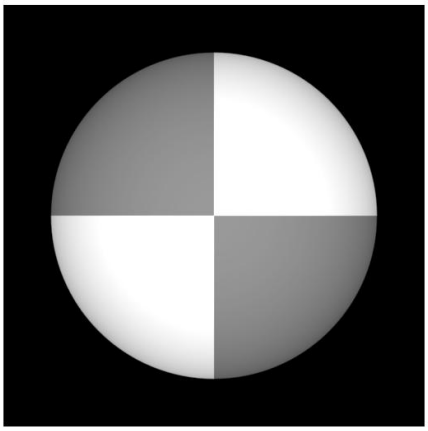


Figure 1: Albedo image of SphereGray5

Estimating the albedo and surface normal requires the number of pictures to be 4 or more, such that the least squares solution can be used. This is given by the fact that for each point in the image, we need to solve the equation $\mathbf{i}(x, y) = \mathcal{V}\,\mathbf{g}(x, y)$.

In photometric stereo, the position of the source of light determines regions of the image to be in shadow (partial or complete shadowing). This creates a few problems in determining the albedo and surface normal, which can be solved through a mathematical trick. The idea is to create a matrix from the image vector and multiply both sides of the above equation by this matrix. Henceforth, the above equation can be rewritten as $\mathcal{I}\,\mathbf{i}(x, y) = \mathcal{I}\,\mathcal{V}\,\mathbf{g}(x, y)$, where $\mathcal{I}$ is a diagonal matrix with all the $\mathbf{i}(x, y)$ vectors. Such operation zeroes out the contributions from shadowed region on the right side of the equation. The shadow trick is more necessary to be used in the case of 5 images, because when there are fewer images, some points will not be in shadow in few instances. However, when more images are used, a point in the image is lightened up in more instances, leading to better final results.

1.2 Question - 2

On the **GrayScale5** set of images, the test shows 2,948 outlier points using the 0.005 threshold. Visualising the results with the *show_results* function, we observe that outliers are present on the edge of the sphere (see upper left image in figure 2). A possible mathematical explanation for this could be that the edges of the object (in this case, the circle defining the sphere contour) lead to the second derivative to have 0 values.

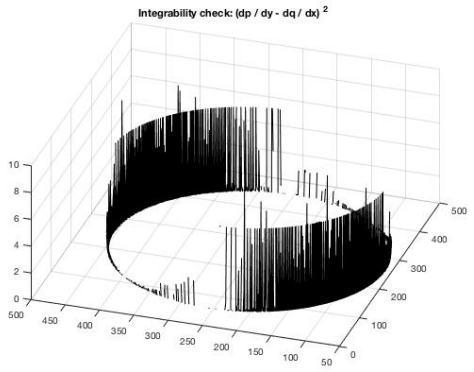


Figure 2: Results of integrability test for GrayScale5

Regarding the number of images used, we observe a negative relationship between the number of images used and the number of outliers found by our test. As we increase the number of input images, the number of outliers decrease as well. The results, presented in table 1, prove the negative correlation. This implies that increasing the number of images lead to a better analysis and thus a better understanding of the edges of the object being photographed.

Number of input images	Number of detected outliers
5	2,948
15	2,668
20	2,452
25	2,231

Table 1: Relationship between input images and outliers

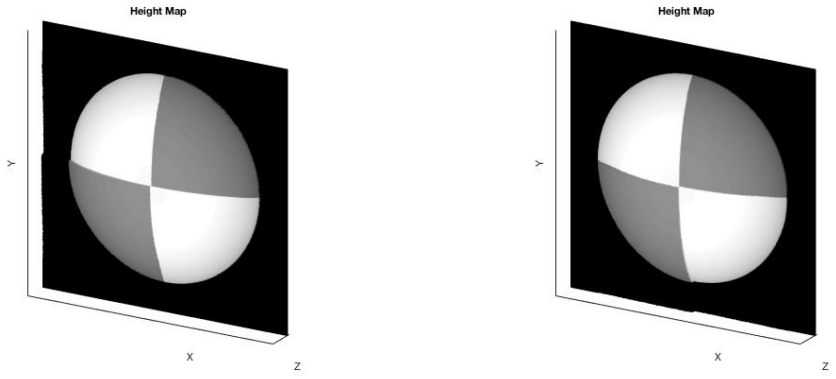
1.3 Question - 3

The results of the two integration paths (column-major and row-major) can be visualised in figure 3, where the left plot is the column-major, while the right one is the row-major. The main difference between them can be seen in a small dent in the figure (object and black background): the column-major solution has this rupture on the Y-axis (left side of the plot), whereas the row-major has it on the X-axis (bottom side of the plot). One possible explanation for this behaviour of the plots is the presence of outliers previously found on the edge of the sphere.

The averaged height model of the column and row integration parts presents a graph with dent on both axis. However, these features are less prominent due than either of the two previous height maps. The resulting plot is shown in figure 4.

1.4 Question - 4

Running this algorithm on **MonkeyGray**, we get the results displayed in figure 5. The many errors from the integrability test are due to the complicated shape of the monkey and the presence of many defining edges. These edges delimit the eyes, nose, ear, mouth, etc. Furthermore, the shading of different parts of the picture varies a lot from picture to picture.



(a) Column-major integration **(b)** Row-major integration

Figure 3: Column-major vs row-major path integration

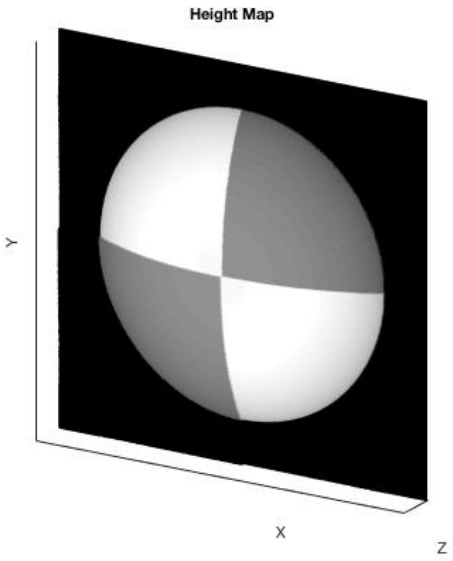


Figure 4: Average surface height map

1.5 Question - 6

The results of the Yale Face model (figure 6) present the three different integration paths applied to the images with faces. It can be observed that the row-major integration path creates a huge distortion of the face on the X-axis at the middle of the plot, while the column-major one creates some distortion at the top on the Y-axis. The averaged path corrects the former problem, but it does not entirely correct the latter one. The Yale Face images violate the assumptions of the photometric stereo algorithm because of the presence of images which present odd characteristics: *yaleB02_P00A+025E+00.pgm* and *yaleB02_P00A+020E-10.pgm*. These two images are somewhat different from the others, having patterns of lines which create weird shadings and textures. Removing the problematic images lead to slightly better results.

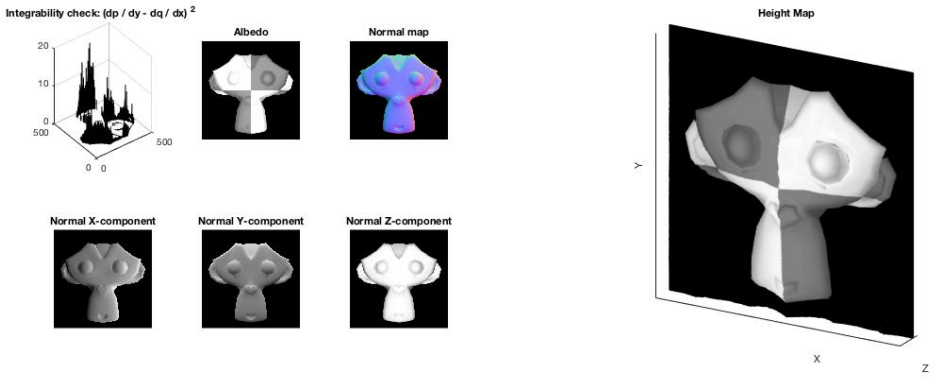
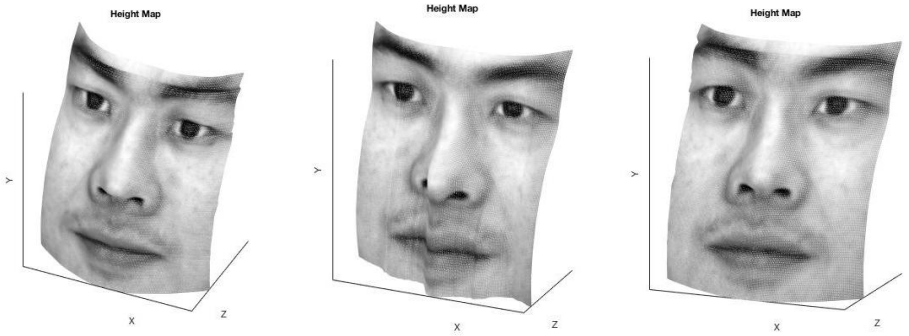


Figure 5: MonkeyGray results



(a) Column **(b)** Row **(c)** Average

Figure 6: Yale Face model integration paths

2 Colour Spaces

2.1 RGB Colour Model

The main reason behind using RGB as the basis in digital cameras and photography is the idea of additive versus subtractive colours. On the one hand, additive colours are emitting elements; the image gets brighter as you add more of them. On the other hand, subtractive colours are reflecting elements; adding more creates a darker colour. From the beginning, the computer monitors were seen as emitting objects. So the RGB colour model, made of three additive colours, became the basis for them and for all digital screens around the world. In contrast, the subtractive colours are used in printing.

Traditionally, the cameras have been built using the principles of coloured phosphors that were also used in TVs. At the "heart" of the digital cameras stands the imaging sensor which is an analogous component. This sensor collects light photons in its photosites, and electrical charges are then produced. These pixel photosites measure the brightness of the red, green and blue light using spectral response functions. Finally, the image processor of the camera interpolates colour data to assign full colour values to all pixels.

2.2 Colour Space Properties

Opponent Colour Space

Because perception of colour is not always best seen in RGB, the opponent colour space presents a different point of view on colours, based on the opponent colour theory. It is composed of three

components: a luminance component, a red-green channel, and a blue-yellow channel. An image in opponent colour space, together with its components can be seen in figure 7.

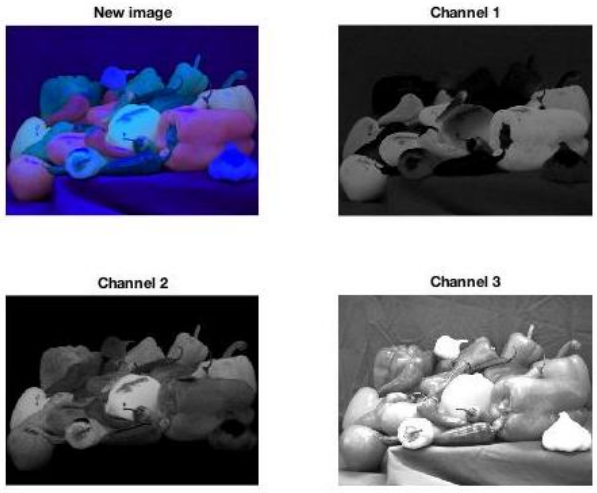


Figure 7: Oponnent colour space

Normalized RGB Colour Space

The normalied RGB (rgb) colour space is used as an effective way of correcting an image of the distortions caused by lights and shadows. The idea behind the normalized RGB space is to take each pixel's value and divide it by the sum of the values over all channels, such that $R' = R/(R+G+B)$, $G' = G/(R+G+B)$ and $B' = B/(R+G+B)$. A transformation into the normalized RGB space is presented in figure 8.

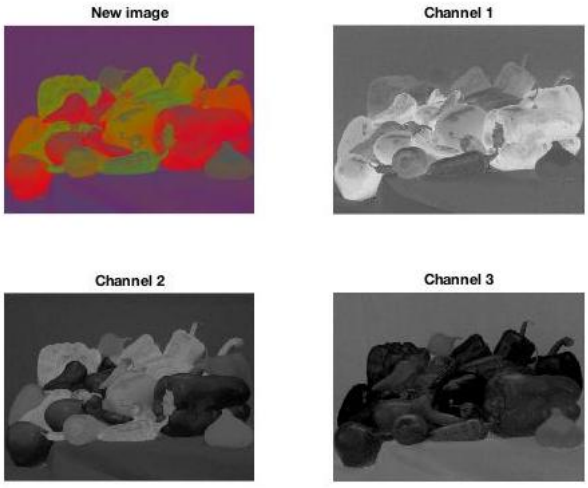


Figure 8: Normalized RGB colour space

HSV Colour Space

HSV stands for hue, saturation and value, and it is a colour space that describes hues (colours) in terms of saturation and the value of brightness. The important difference of HSV compared to RGB is that image intensity is separated from the colour information (luma and chroma). In the computer vision context, it is important to have this separation because this adds robustness to lightning changes or in case of shadow removal. The resulting HSV transformed image is shown in figure 9.

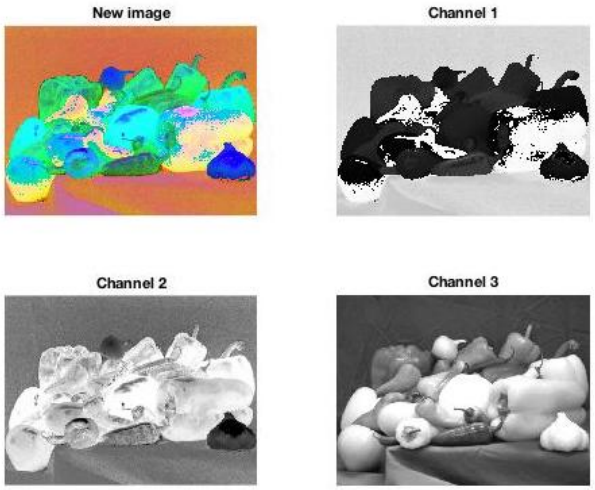


Figure 9: HSV colour space

YCbCr Colour Space

In the YCbCr colour space, the Y component addresses the brightness of the colour (luma), which is the light intensity. The other two components, Cb and Cr are the blue component and red component relative to the green component respectively. Figure 10 below presents the resulting YCbCr image and its separate components.

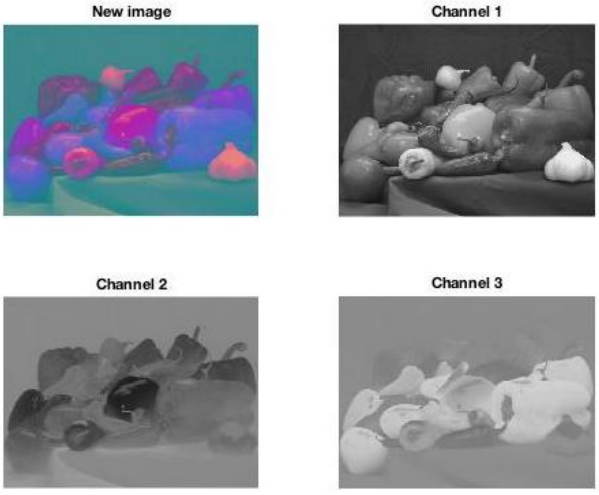


Figure 10: YCbCr colour space

Grayscale

The greyscale methods convert a three-channel image, defined by the triplets (R, G, B), to a single channel image, defined by a single grayscale value. Converting RGB images to grayscale has the main advantage of reducing the inherent complexity created by colours. Images in grayscale can be easily studied in image processing to understand more about contrast, shapes, edges, textures, etc, while not addressing the issue of colours. Another advantage of the grayscale images might also be the lower storage requirements of a single channel instead of three. There are multiple methods of transforming an image to grayscale. In the *lightness method*, the maximum and the minimum value of the triplet (R, G, B) are summed up and divided by two, thus averaging over the most and least prominent colours. The *average method* just takes the average of the individual R, G and B values, while the *luminosity method* averages the three values, but uses a weight for each colour: 0.21 for red, 0.72 for green and 0.07 for blue. The results of each method can be seen in figure 11.

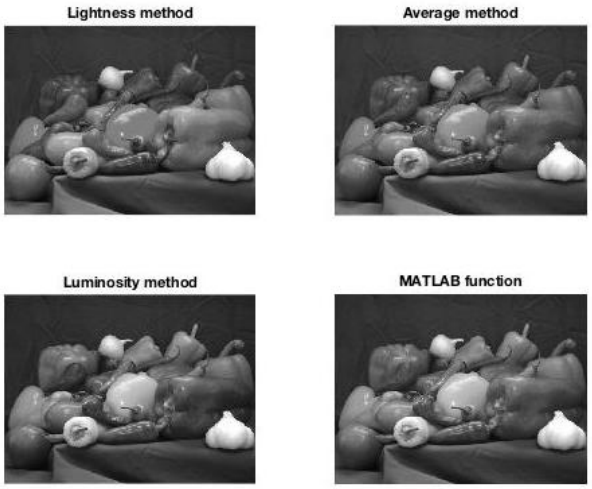


Figure 11: Grayscale methods

2.3 More on Colour Spaces

CMYK

The CMYK colour space is widely used in the printing process and uses four inks: cyan, magenta, yellow and black. The main idea behind using CMYK in printing is that of masking (partially or entirely) colours on a white canvas/background. Because of the subtractive nature of these ink, they are applied in such a way which impedes light to be fully reflected, creating the desired image.

3 Intrinsic Image Decomposition

3.1 Question - 1

Two other components images can be decomposed in are structure-texture and signal-noise. Structure-texture: Given an image f, it is possible to split it into two components u and v, u containing the geometric information and v the texture information. The idea is that an image is composed of a structural part, corresponding to the main large objects in the image, and a textural part, containing fine scale-details, usually with some periodicity and oscillatory nature. [1] Signal-noise: Given an image f, it is possible to split it into two components u and v, u containing the original image and v the noise. The goal is to estimate the original image by removing the noise from a noise-contaminated version. It has been proved that the use of wavelets successfully removes noise while preserving the signal characteristics, regardless of its frequency content. [2]

3.2 Question - 2

It is very hard to get the ground-truth from real images. If you want to use real image, something like crowd sourcing is needed to find the ground-truth. This will become very expensive for large data sets, which are often needed for machine learning tasks. With synthetic images, you know the ground-truth because the images are created using these values. [3]

3.3 Question - 3

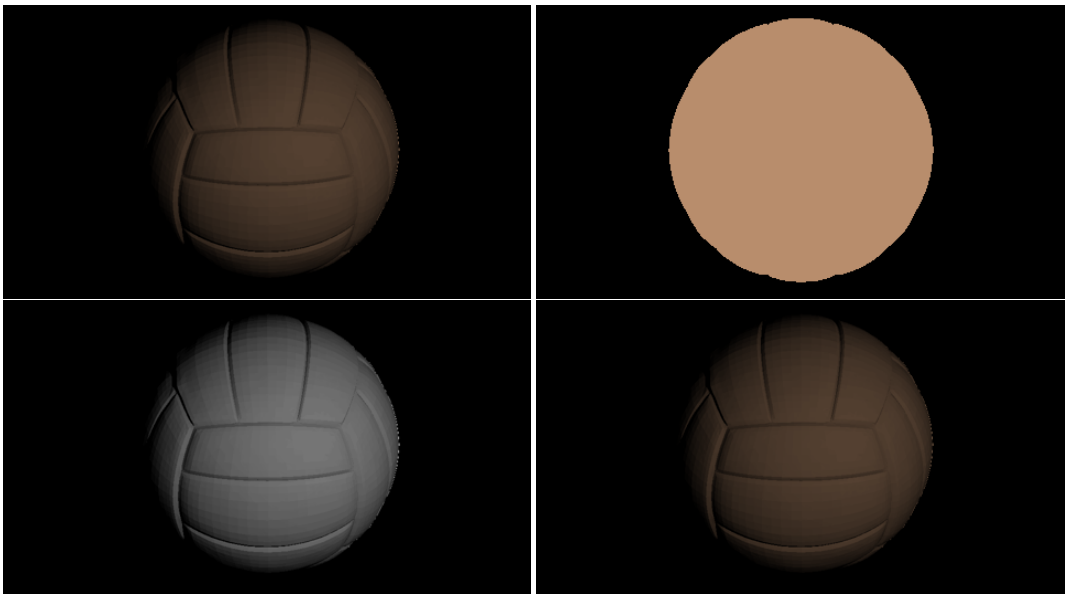


Figure 12: [1] original image, [2] reflectance, [3] shading, [4] reconstructed image

3.4 Question - 4

The true material colour can easily be found by in this picture by looking at the reflectance. The reflectance matrix shows the colour and is independent of confounding illumination effects. Looking in the matrix, we find RGB = (184, 141, 108).

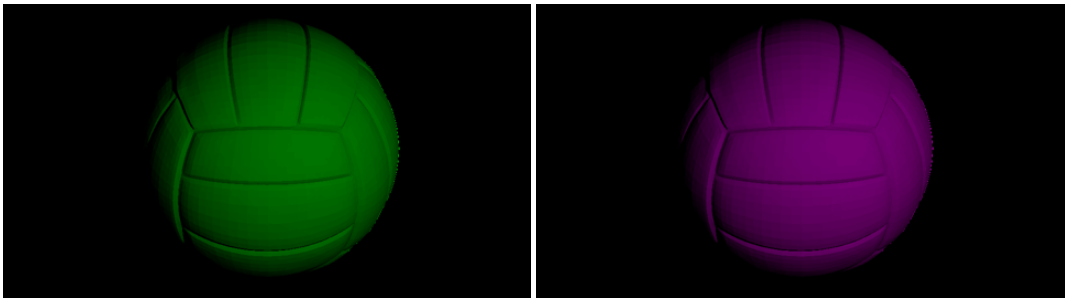


Figure 13: [1] recoloured with green, [2] recoloured with magenta

Although the object is recoloured with pure colours, the reconstructed images won't have a uniform colour. This is because the shading doesn't have uniform values. As can be seen from the formula:

$$I(\vec{x}) = R(\vec{x})xS(\vec{x})$$

With $I(\vec{x})$ the observed image, $R(\vec{x})$ the reflectance and $S(\vec{x})$ the shading, every pixel in the reconstructed images can be multiplied with a different value because of the shading, causing a uneven reconstructed image.

4 Colour Constancy

4.1 Question - 1

The function using the Grey-World algorithm calculates the mean of the three channels. Assuming that under a white light source, the average colour in a scene is achromatic (grey, [128, 128, 128), it

calculates the correction needed for the three channels. This correction is multiplied with the channels in order to get the white balanced image.

Figure 14: [1] original image, [2] white balanced image using Grey-World

In this case the Grey-World algorithm was able to removed the reddish colour cast on the image. The algorithm might fail when the assumption of a achromatic scene is invalid. For example a close-up image of a green apple taken with a white light source. The Grey-World algorithm will boost the red and blue's, while decreasing the green's. This wasn't needed since the image was already composed of the right colours.

Figure 15: [1] original image, [2] white balanced image using Grey-World

Two other colour constancy algorithms proposed in the literature are the white patch algorithm and singlescale retinex (SSR). white patch assumption: the white patch approach uses the maximum value from the three channels to calculate the correction. It makes the assumption that the maximum should be 255 in a white balanced image. This approach is similar to how the human visual system does colour correction. singlescale retinex (SSR): Proposed by Edwin Land in 1986. An method for bridging the gap between images and the human observation of scenes. The design of the singlescale retinex consists of: the choice of a surround function; the placement of the log function and the final signal processing prior to display. [4]

References

[1] Jean-François Aujol, Guy Gilboa, Tony Chan, and Stanley Osher. Structure-texture image decomposition—modeling, algorithms, and parameter selection. *International journal of computer vision*, 67(1):111–136, 2006.

[2] Rupinderpal Kaur and Rajneet Kaur. Survey of de-noising methods using filters and fast wavelet transform. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(2), 2013.

[3] Sean Bell, Kavita Bala, and Noah Snavely. Intrinsic images in the wild. *ACM Transactions on Graphics (TOG)*, 33(4):159, 2014.

[4] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Retinex image processing: Improved fidelity to direct visual observation. In *Color and Imaging Conference*, volume 1996, pages 124–125. Society for Imaging Science and Technology, 1996.