Sampling Strategies

Jurriaan Heuberger, Wouter Pasman

Sampling strategies

The data fitting experiments have shown that the exclusion of the extreme angles from the data fitting actually increases the accuracy of the fitting when measured for this limited range. However, we would also like to focus the sample positions around the ideal reflection angle. This is where the largest variation in the signal occurs, and we want to accurately capture the variation. By placing more emphasis on the specular peak, we can model it more accurately using fewer sample points.

Specifically we want to vary the sampling positions around the ideal reflection direction over the hemisphere of outgoing directions for all angles θ_o and φ_o. We now move from the rectangular space of (θ_o, φ_o) to the also rectangular space $(\Delta\theta, \Delta\varphi)$ where $\theta_o = \theta_r + \Delta\theta$, and $\varphi_o = \varphi_r + \Delta\varphi$. We try to divide this space in such a way that the highest density of sample points is around the point (0, 0).

Sampling setups

If we use a simple regular rectangular grid type setup, we can divide the sampling hemisphere into equal sections. This gives a regular sampling of the hemisphere into fixed blocks of a certain size x. However, it is also possible to adjust the grid distances in such a way that we get a grid that is more densely sampled in the center as shown in figure 1. This irregularly distanced rectangular grid places more emphasis on the center while maintaining a very regular nature.

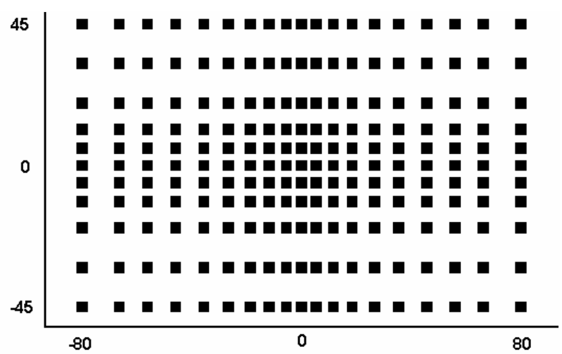

Figure 1: Irregular rectangular grid

This setup has the advantage that it is very easy to iterate through all the positions. To get from one position to the next, only 1 of the two variable directions needs to be adjusted. The other direction remains fixed. Another advantage is that it is very easy to alter the grid in order to increase or decrease the sample density in the horizontal or vertical direction.

A disadvantage is that this setup requires a quite large number of sample positions. Especially in the farther regions of the sample space we might have more sample points than we really need. The density of the $\Delta\varphi$ parameter division is the same at + 45 and -45 for $\Delta\theta$, as it is around 0. These points however are far less relevant. The same is true for the division of $\Delta\theta$ at the extreme reaches of $\Delta\varphi$.

One way to avoid this problem is to reduce the density of the grid at certain threshold values. Doing this for one parameter already creates a reduction. And doing it for both parameters reduces it further.

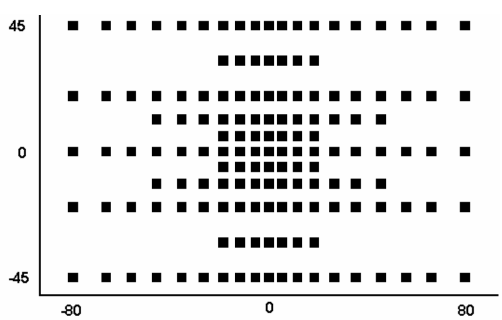

Figure 2: Vertical density reduced after 3 and 6 steps

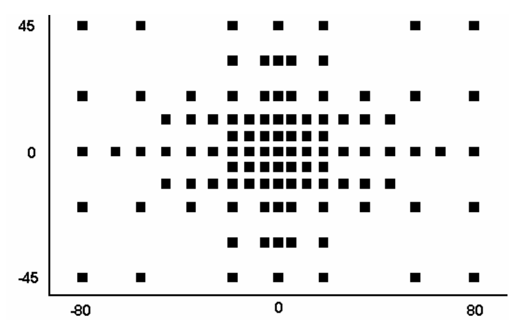

Figure 3: Vertical density reduced after 3 and 6 steps, Horizontal after 2 and 4 steps

Another possible setup is to use an elliptical setup, in which the sample space is subdivided into a number of concentric ellipses figure 4. Each ellipse contains a fixed number of sample positions at equal distances from each other. This setup automatically guarantees that there will be more sample positions around the 0° direction. To further influence the density of the sample points it is possible to use different distances between the ellipses.

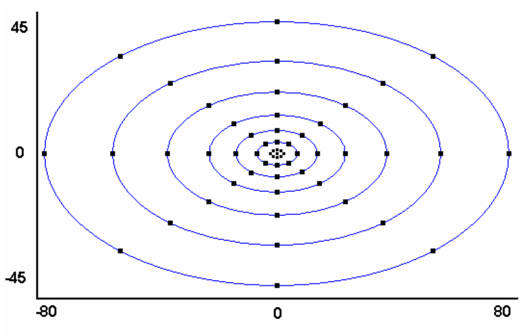

Figure 4: Elliptical sampling

An advantage of this setup is that the number of sample positions is much lower than the rectangular grid, while still maintaining a relatively good spread of points around the sample space. Unfortunately every position has 2 degrees of freedom. This means that almost every position change would require movement in 2 directions across the hemisphere, and that means more measuring is needed.
Ideally we would like the lower density of the elliptical setup, with the regularity of the irregular rectangular grid. A possible solution is replacing the ellipses with rectangles. Each concentric rectangle has an equal number of points on it. These may not be equidistant anymore, which gives a mechanism with which to influence the division of points in both directions. Once again it is possible to further influence the density of samples by increasing the size of the consecutive concentric rectangles exponentially.

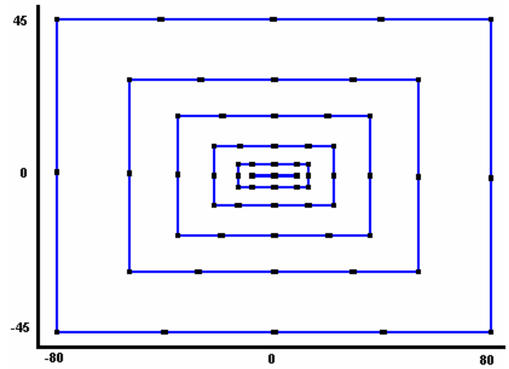

Figure 5: Rectangular sampling

A possible problem that may arise is that because of the regular nature of the sampling positions, aliasing effects start to occur. A possible solution to this for the elliptical sampling is jittering. By randomizing the initial position so that the rotation of the points on each ellipse is different, aliasing effects should be reduced. Jittering to avoid aliasing can also be performed on the rectangular sampling. By altering the precise position of the points on each side of the rectangle, a randomness is introduced which should limit the aliasing effects.

Analysis

We can perform the data fitting with these reduced sets of data points, and examine how well the model fits to the measurements. In our previous work we already determined that for our purposes the Lafortune BRDF model is best suited. Specifically a 2 lobe Lafortune model outperforms a single lobe model. Therefore we only fit the reduced data sets to the 2 lobe Lafortune model.
Just like before we want to quantify the goodness of fit, and we use the same measures as before. We quantify how well the fitting went by calculating a scaled signal to noise ratio, SNR', of the model to a regular 1° sampled data set with θ_o limited to 60°. We compare this value with the values obtained in the previous experiment to determine how good the fitting was. Comparisons are of course twofold. First of all we compare the value of the SNR' to determine which data set results in the best fitting. Second of all, a comparison of the number of data points can indicate whether a set achieves a good fit with lower number of data points. A comparison of both these values is used to determine which data set gives the best results.

Setup

In order to test each of the three setups listed above we created a number of data sets based on the different sampling approaches. For the reduced sets we start with a full manual set as used in the original experiment. We then reduce it somewhat to make a second set and more for a third set. This is done for 3 different densities, giving a total of 9 different sets of data samples.
For the other two sample setups, we use 4 different rectangle or elliptical densities. This gives a total of 17 data sets as indicated in table 1. The full set shows the angles at which measurements are taken, and

the reduced set shows the index to which the parameters are used. So for the first set an index of 16 for a value indicates it is used over the entire range of $\nabla\varphi$ values, while a value of 7 means after 10° this values is no longer used. For the elliptical sets, the values indicate the left and top values of concentric ellipses with the origin at 0°. For the rectangles these values are the same, but using concentric rectangles instead. Each of these rectangles or ellipses contain 8 points which are equally spaced on the ellipses and are the corners and midpoints of the edges for the rectangles.

	Parameter $\Delta\theta$	Parameter $\Delta\varphi$
Full set I	0, 1, 2, 3, 5, 7, 10, 15, 25, 35, 45, 60	0, 1, 2, 3, 5, 7, 10, 15, 25, 35, 45, 60, 90, 115, 140, 170
Reduc. 1 set I	16, 5, 7, 16, 16, 9, 16, 16, 16, 16,16, 16	12, 5, 7, 12, 12, 9, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12
Reduc. 2 set I	16, 5, 7, 16, 16, 9, 16, 9, 5, 9,16, 16	12, 5, 7, 12, 12, 9, 12, 9, 5, 9, 12, 12, 12, 9, 12, 5
Full set II	0, 1, 2, 5, 10, 17, 25, 40, 60	0, 1, 2, 5, 10, 17, 25, 40, 60, 90, 125, 155
Reduc. 1 set II	12, 4, 12, 5, 12, 5, 12, 12, 12	9, 4, 9, 5, 9, 5, 9, 9, 9, 9, 9, 9
Reduc. 2 set II	12, 4, 5, 5, 12, 5, 5, 12, 12	9, 4, 5, 5, 9, 5, 5, 9, 9, 9, 4 ,9
Full set III	0, 1, 3, 7 ,15, 30, 45	0, 1, 3, 7 ,15, 35, 60, 90, 155
Reduc. 1 set III	9, 4, 9, 5, 9, 9, 9	7, 4, 7, 5, 7, 7, 7, 7, 6
Reduc. 2 set III	9, 4, 5, 5, 9, 7, 8	7, 4, 5, 5, 7, 6, 7, 7, 5
Rect/Elliptical I	1, 2, 3, 5, 7, 10, 15, 30, 45, 60, 75	1, 2, 3, 7, 15, 30, 45, 60, 90, 135, 155
Rect/Elliptical II	1, 2, 3, 7, 10, 15, 30, 45, 60, 75	1, 3, 7, 15, 30, 45, 60, 90, 135, 155
Rect/Elliptical III	1, 3, 7, 15, 30, 45, 60, 75	3, 7, 15, 30, 45, 60, 90, 135
Rect/Elliptical IV	1, 3, 7, 15, 30, 45	3, 7, 15, 45, 60, 90

These data sets are calculated for different θ_i values as well. First they are calculated for a regular 10° sampling from 0° to 80°, then for a different set of incoming angles based on the principle that paintings will often be lit from the side more than they will be lit from the front. This second set of angles consists of [85, 80, 75, 65, 55, 45, 30, 15]. Because these angles are further towards the extreme angles (and the ideal reflection will be as well) the number of data points within the range of [0° 60°], will be smaller than with the regular sampling.

The initial results of the investigation led us to make some additional sets of data. The overlap between the different types of sets in terms of the number of sample points was minimal, so in order to achieve a greater overlap, more expansive sets indicated in table 2, were used. These were used with the standard 8 points per ellipse/rectangle, and also with 12 points per ellipse/rectangle, with equidistant angles for the ellipses, and 5 points per φ edge for the rectangles.

	Parameter $\Delta\theta$	Parameter $\Delta\varphi$
Rect/Elliptical V/VII	1, 2, 3, 4, 5, 7, 10, 15, 20, 25, 30, 45, 60, 75	1, 2, 3, 5, 7, 10, 15, 25, 30, 45, 60, 90, 135, 155
Rect/Elliptical VI/VIII	1, 2, 3, 4, 5, 7, 10, 15, 20, 25, 30, 35, 45, 60, 75	1, 2, 3, 5, 7, 10, 15, 25, 30, 45, 60, 85, 90, 135, 155

These extra sets were only made for second set of θ_i values. In total this gives 42 different data sets which were fitted to the data of metallic blue paint from the MIT/MERL database as was done before.

Results

After the data fitting was performed the following SNR' values were obtained.

	Regular sampled θ_i		Extreme θ_i angle set	
	Nr Samples	SNR'	Nr Samples	SNR'
Full set I	14756	17.54	10664	16.51
Reduc. 1 set I	11196	17.59	8120	16.64
Reduc. 2 set I	8332	17.46	5768	16.04
Full set II	8188	17.49	5704	16.36
Reduc. 1 set II	5324	17.37	3720	16.33
Reduc. 2 set II	3332	16.79	2440	16.10
Full set III	4420	16.58	3536	16.47
Reduc. 1 set III	3504	17.04	2584	16.67
Reduc. 2 set III	2784	16.65	2000	16.79
Elliptical I	1888	12.44	1324	16.70
Elliptical II	1688	12.58	1196	16.81
Elliptical III	1320	12.12	932	16.83
Elliptical IV	1112	16.02	784	16.52
Elliptical V			4556	17.18
Elliptical VI			3336	16.25
Elliptical VII			2256	15.41
Elliptical VIII			2108	15.62
Rectangular I	1904	12.83	1316	16.31
Rectangular II	1704	12.89	1188	16.48
Rectangular III	1292	15.71	932	16.65
Rectangular IV	1120	16.80	784	15.89
Rectangular V			4520	17.07
Rectangular VI			3320	16.49
Rectangular VII			2256	15.84
Rectangular VIII			2108	15.96

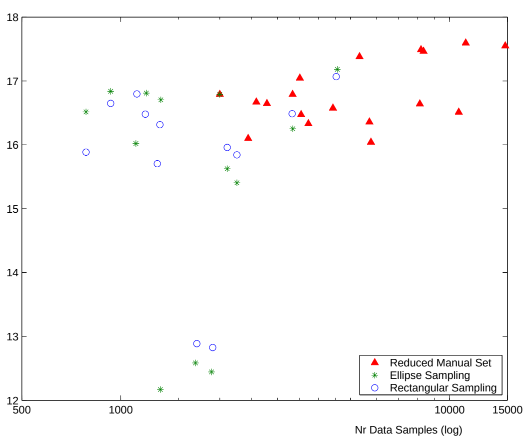

Figure 6: SNR' values of reduced sampling

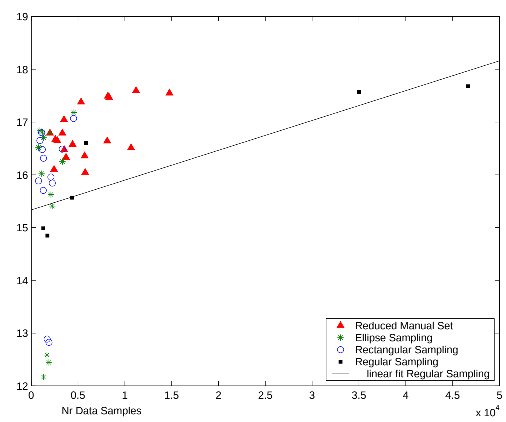

Figure 7: SNR' compared to regular sampled output limited data

These can be plotted according to the number of sample points as indicated in figure 6. If these values are compared to the SNR' of the original regular sampled output range limited data sets from the original experiment, we see a higher SNR' value for comparatively far fewer sample points.

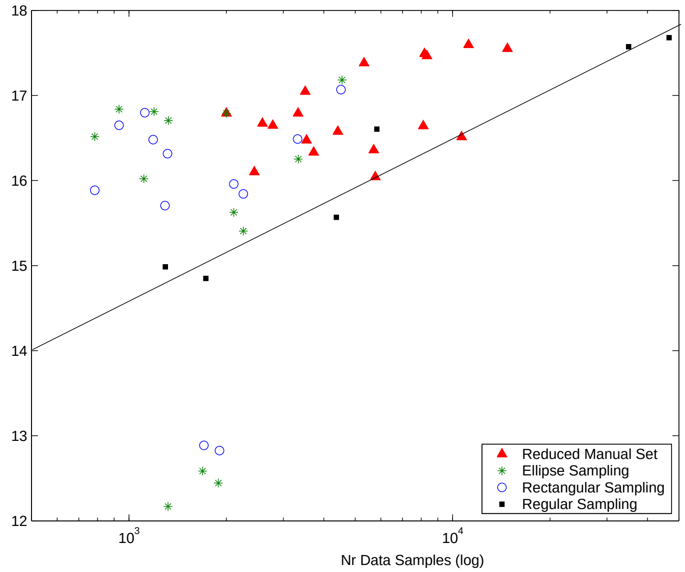

Figure 8: SNR' compared to nr of samples in log space.

If the X axis is plotted in log space, the linear coherence of the original SNR' values becomes stronger. What is clear is all the pictures is that though there are some differences in the value, in general all values are above line indicating the linear fit of the original data, indicating a better general fit.

Conclusion

The graphs show that it is possible to get relatively very good SNR' values with comparatively few data points. Almost all the values obtained are above the line indicating the regularly sampled sets from the previous experiment. The sets with more than 4000 sample points, however, have a noticeably higher SNR' value than the sets with fewer sample points. The difference is about 0.5 to 1 unit, which in log space is significant. It is also noticeable that the use of the alternate set of θ_i values results in a lower SNR' value. However even the sets with 2000 or less (less than 50% of the above mentioned sets) have values in the range 16-17. This is a significant improvement over values in previous experiments.

Discussion

After the completion of the experiment there are still a number of aspects that are open to discussion. First of all, there are certain sampling sets that contain sets which have a substantially lower SNR' values. These sets are specifically low in sample density and contain the incoming light direction 0°. The inherent problem of this light direction is that camera and light direction of the ideal reflection direction are the same. This leads to measurement problems as the light interferes with the camera or vice versa. By replacing the value 0° in those sets with a small value such as 5°, results in fittings with SNR' values in the range 16.8-17.3. Excluding the 0 value seems to have a beneficial effect on the fitting accuracy.

Another point of discussion is how well the model could actually fit to the measured data. Currently we have no SNR' values greater that 18. This could be because we haven't managed to fit well enough because we are not using the right points (or enough points), but there is of course also a limit to how well the model will ever fit to the measured data. It is quite possible that this limit is very near in which case taking more sample points to perform a fitting, has no discernable advantage.