The cost function and quality criterion used for distortion estimation and correction is the root mean square distance of sets of points from best-fit straight lines. This is plausible indeed. One critical issue however is that this measure is not by default invariant to the degrees of freedom that are unconstrained by distortion correction. Depending on the distortion model, the distortion correction is defined up to a 2D homography (for a basic distortion model, such as radial distortion with known distortion center, the result is more constrained though). There are two issues with this: [1] to avoid a trivial solution, one needs to take care of this; in the case of the model and method of chapter 6, you do this by imposing values on particular coefficients of the distortion model used. [2] most importantly: how to objectively compare distortion correction results produced by two different models if one model is defined up to a 2D homography and the other not? This is not addressed in the thesis, but is essential. Please explain if you have done anything about this.

It is true that the measurement of "straightness" is not invariant to a 2D homography. This measurement can be enlarged or reduced by the scale introduced by the homography. This is related to the two issues you mentioned in your email.

One issue is to avoid the trivial solution by imposing values on particular coefficients of the distortion model used. In Chapter 6, coefficients of order-0 terms and order-1 terms of the polynomial model are set such that the correction is close to identity at the center of image. Namely, for the x-component of the polynomial model, the coefficient of the term $(x - x_c)$ is set to 1, the coefficient of the term of $(y-y_c)$ is set to 0, and the coefficient of order-0 term is set to 0; for the y-component of the polynomial model, the coefficient of the term $(x-x_c)$ is set to 0, and the coefficient of the term of $(y-y_c)$ is set to 1, and the coefficient of order-0 term is set to 0. This setting up avoids the trivial solution and also fixes the scale introduced to the measurement of "straightness" (apply another homography on the corrected points will change these particular coefficients). This setting up allows us to compare different polynomial models objectively.

Another issue is how to compare the correction results of two different models if one model is defined up to a 2D homography and the other not. This point is very important and is not sufficiently discussed in the thesis. I think there are two strategies: one is to estimate the homography between the points corrected by two models (or estimate the homography between the lines fit to the points corrected by two models) and compensate the homography by applying it on one model; another strategy is to find two homographies for two models such that the two models are normalized such that the correction is close to identity at the center of image. This normalization is equivalent to set up the coefficients of order-1 and order-0 terms in the models. A more practical method is to normalize the "straightness" error by the average length of all the corrected lines of different orientation (if we have many lines of different orientations uniformly distributed in the image domain, this normalization captures the scale introduced by the homography).

In the experiments, some particular coefficients of polynomial models are fixed as above. The coefficients of order-1 terms of the distortion model used in Lavest's method are also fixed to be 1. And this is also the case for the radial model used in Alvarez's method. For the non-parametric method based on the textured pattern, the photo is taken such that the whole camera captor is covered by the whole pattern. This gives a weak homography which does not introduce a big factor to reduce or enlarge the correction error. And for the comparison, the "straightness" error is not the only criteria, we are also interested in the shape (or global tendency) of the curve of "straightness" error.

The points you indicated are really important and I agree that I have not sufficiently discussed in the thesis. For a general measurement of the distortion correction precision, a criteria invariant to 2D homography is indispensable. In our experiments and comparisons, we set the particular coefficients in parametric models (and take photos with the whole camera captor covered by the whole pattern for the non-parametric textured pattern-based method) to limit the influence of the homography.

Just for curiosity: did you have access to the original code by Lavest for the experiments?

For the code of Lavest, he kindly gave it to us during our visit to his lab. We did the following modifications to his code: 1) the correction of bias by using the ellipse centers (what we really need is the projection of the circle centers on the pattern, which are different from the ellipse centers); 2) the calibration pattern containing more circles (180 circles) is used.

The above modifications are small and give the results similar to the original code.

In the experiments, if the threshold of re-projection error is set to be too low, the points whose re-projection error bigger than this threshold will be removed and the algorithm will be iterated until all the points have the re-projection error smaller than the threshold. To avoid the problem, we did not set a too small threshold (about 0.1 pixels in the experiments) such that no point will be removed.

Table 5.4: is there any explanation why for several columns, entries in the first five rows, are identical?

For Table 5.4, the first five rows are identical. This is because a tangential distortion is added. The first five rows represent radial model, division model and FOV model. These models are radial symmetric and are not capable to correct tangential distortion. These three models are almost self-consistent to each other, so the error shown in Table 5.4 (the first five rows) are identical (only the order of magnitude of the value is shown in Table 5.4).

Section 5.5: it is not entirely clear with which setup the distortion was estimated with the different models, i.e. which type of images and what number of images were used for the different models. Could you please explain this briefly?

In section 5.5, we use the straight lines to verify the correction performance of different models. First, two photos of a digital textured pattern are taken and the loop validation in Chapter 4 is used to find the "inliner" matchings between the digital pattern and one of its two photos. These matchings are used to fit different distortion models and the average/max fitting error is presented in Table 5.5. Second, one photo of straight lines is also taken. This photo is corrected by different fit models. Then we can measure the average/max straightness error of the corrected lines in Table 5.6.

Table 5.5: in some instances, the fitting error increases slightly with increasing model order. Why is this the case? (I had the impression that higher order models are initialized with the results of the preceding lower order model).

Table 5.5 shows the average/maximal value of fitting error. Since we minimize the average error (not the maximal error), in some instances, the maximal error increases with the increase of order. I also observed that the average error sometimes increases very slightly (0.01 pixel for radial model, division model and FOV model). This should not happen because higher order models are initialized with the results of the preceding lower order model. I think that this is due to the non-linear minimization which I do not master in.

Table 5.6: same question as for Table 5.5. Also, it is not clear if for these results, the distortion center was estimated or not for the models which do have a distortion center. If not, how was the center set? Further, what is meant by the order of the FOV model?

In Table 5.6, the fit distortion models are used to correct the distorted lines. Since the fit distortion models are estimated by fitting the matchings between the digital pattern and one of its photos, the increase of order of models decrease the average fitting error as shown in Table 5.5. But the fitting error is not directly related to the straightness error in Table 5.6. Only when the fit models can well approximate the deformation between the digital pattern and its photo (lens distortion + homography), the straight lines can be well corrected and the straightness error decreases with the model order. This is the case of polynomial model and rational model. But for the radial model, division model and FOV model, since they are not capable to approximate the the deformation between the digital pattern and its photo (lens distortion + homography) (Table 5.5 shows it), the distortion lines are not well corrected. So the increase of model order does not necessarily lead to the better correction of the distortion lines. In addition, the algorithm (LSD) used to detect the lines can miss some badly corrected lines.

Table 5.5 and 5.6 reflect the correction precision of different modes in two aspects: fitting error and straightness error. In the experiments, the distortion center are estimated for the models which do have a distortion center (radial, division and FOV model). The distortion center is initialized by the center of image and then refined by non-linear minimization. The original FOV model proposed by Devernay and Faugeras is order 1. The order of FOV can be extended by adding the radial terms of the other orders.

Page 112: as for the general case of unknown orientation of lines, it would be good to completely summarize the estimation approach:

which are exactly the unknowns (i.e. are line orientations explicitly estimated or only implicitly as functions of other variables), are numerical or analytical derivatives used?

As for the general case of unknown orientation of lines, the minimization problem becomes non-linear. The orientations are parametrized by the parameters of polynomial model. So the only unknowns are the parameters of polynomial models. The analytical derivatives are used in the Levenberg-Marquardt algorithm.

Page 114: Concerning the second strategy. Just for clarity: what exactly is done by "iterating the linear method"? Same page, fourth strategy: what exactly is meant by "iterative linear minimization"?

The second strategy is to iterate the linear method. The linear method requires the orientations of the lines. We first compute the linear regression lines from the distorted points. The orientations of the linear regression lines are available and then we can apply the linear method to estimate the parameters of the polynomial model. Once the parameters of the polynomial model are estimated, we correct the distorted points and compute again the orientations of the linear regression lines from the corrected points. The new orientations allow us to apply again the linear method to estimate the parameters of polynomial model. This procedures can be iterated until the average error does not decrease significantly (for example, 0.01 pixel in the experiments).

The "iterative linear minimization" is exactly the same thing as the second strategy except that in the second strategy, the initial orientations are obtained from the linear regression lines of the distorted points, while in the fourth strategy, the initial orientations are obtained from the linear regression lines of the points corrected by incremental LM minimization.

Page 121 and following: In figure 6.3 and similar ones, were the images used for evaluation, also used for the actual distortion estimation?

The image used for evaluation is not used in the distortion estimation. The idea is to first use images of the harp with different orientations to estimate the distortion model, then use another image which is not used in the estimate to evaluate the result.

Page 123: which images were used for calibration using the Lavest method? The same as those shown in figure 4.15? Were they taken at the same time as the images shown in figure 6.5?

The images used for Lavest's method (page 123) look like those shown in Fig. 4.15 (but the experiment on page 123 is different from that in Fig. 4.15). The images used for Lavest's method (page 123) are taken at the same time as the images containing straight lines in Fig. 6.5.

Table 6.7: similar to above for other tables, why does the RMSE

increase sometimes with increasing model order?

The first column of Table 6.7 is not the model order, it is the index of lines. This table shows that with different number of parameters used in the distortion model, the performance of correction is different. This implies that there exists a compensation between the distortion model and the other parameters of camera in the global camera calibration.

Table 6.7 and other: how many different trials did you perform with the different models/methods each, I mean calibration trials with different sets of images?

We did many trials for each model/method, at least five for each, with different sets of images taken in different days. In the thesis, some typical results are shown, but the results on the other sets of images are similar.

The trials are also very related to the development of experimental materials we used, in particular the harp we used. At first, the harp is quickly fabricated with sewing strings. And the photos are taken against the sky to avoid the shadow problem, which is in fact a difficult task. Finally, we used the opaque fishing strings and translucent paper as background to facilitate the experiments. For the non-parametric methods with a textured pattern. We tested the images in the Bordatz textures to compose a pattern which gives a dense SIFT matchings.

Page 145: I'm not sure I understand why the Hartley method can get stuck in a local minimum. As far as I recall, the method only solves linear equation systems.

Page 145, it is true that Hartley's method is linear. It is not appropriate to say that Hartley's method is stuck at a local minima. It is more appropriate to say that Hartley's method uses a metric which does not really reflect the projective distortion introduced by the rectification process.

Page 145: what is meant by "the geometry of the rectified images is not correct"?

Page 145, "the geometry of the rectified images is not correct" means that when the camera motion is forward, the baseline between two cameras is more parallel to the y-axis, so are the epipolar lines. From the geometric viewpoint, to achieve the rectification, the images should be rotated about $90°$. If an algorithm does not rotate the image, we say that it is not geometry correct.

Page 147: As shown, Hartley's method does not rotate any of the two images. This is counterintuitive, since the method's first step, as far as I recall, should rotate at least one of the two images such that the epipolar lines become horizontal.

Page 147, in fact, a variant of Hartley's method was tested in thesis, it is the code of Du Huynh available on: `http://www.csse.uwa.edu.au/~du/`

`Software/Welcome.html`. Compared to Hartley's method, this code has feature that if the original epipolar lines are more vertical than horizontal, it will do the rectification such that the epipoles are sent to $(0,1,0)^T$; while if the original epipolar lines are more horizontal than vertical, it will do the rectification such that the epipoles are sent to $(1,0,0)^T$. So in fact, what we are comparing is the Du Huynh's method, even though it is very similar to Hartley's.

Page 149: Is it possible to run Mallon's method also for the last three datasets?

Page 149, the Mallon's code is not available on line. But we will implement it soon. We want to implement all the image rectification algorithms and make a benchmark. This is the idea of IPOL (Image Processing OnLine `www.ipol.im`), which is the journal publishes relevant image processing and image analysis algorithms, where the algorithm can be tested on line on data sets uploaded by the users.

Page 163 and following: please provide examples of images plus matches that were found as inliers for the estimated homography. In particular, how many inlier matches were obtained for the different methods on the shown examples?

Here I show an example of the affine transformation, represented by the matrix:

$$A = R_1(\psi)T_tR_2(\phi) = \begin{pmatrix} \cos 24^\circ & -\sin 24^\circ \\ \sin 24^\circ & \cos 24^\circ \end{pmatrix} \begin{pmatrix} 2^{4/12} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos 37^\circ & -\sin 37^\circ \\ \sin 37^\circ & \cos 37^\circ \end{pmatrix} \tag{1}$$

Images plus "inliner" matchings in the first ocatve, for Lowe's SIFT method, improved SIFT by cancelling the sub-sampling in scale space and the iterative SIFT by applying the estimated homography, are shown in Fig. 1.

Table 1 shows the number of "inlier" matchings under different transformations in different octaves for Lowe's SIFT. Table 2 shows the number of "inlier" matchings under different transformations in different octaves for improved SIFT by canceling the sub-sampling in scale space. Table 3 shows the number of "inlier" matchings under different transformations in different octaves for iterative SIFT by applying the estimated homography.

Page 164: it is said that for 0 or 90 degrees rotations, results are better than for in-between rotations and that this is due to the imprecision in orientation estimation in SIFT. Would another explanation be that for 0 or 90 degrees rotation, the interpolation required to generate rotated images, introduces less artefacts than for other angles?

Page 164, The explanation that for 0 or 90 degrees rotation, the interpolation required to generate rotated images, introduces less artefacts than for the in-between angles is also plausible. In fact, if the rotation angle is 0 or 90 degrees, no interpolation is needed. So there is no artifacts introduced.

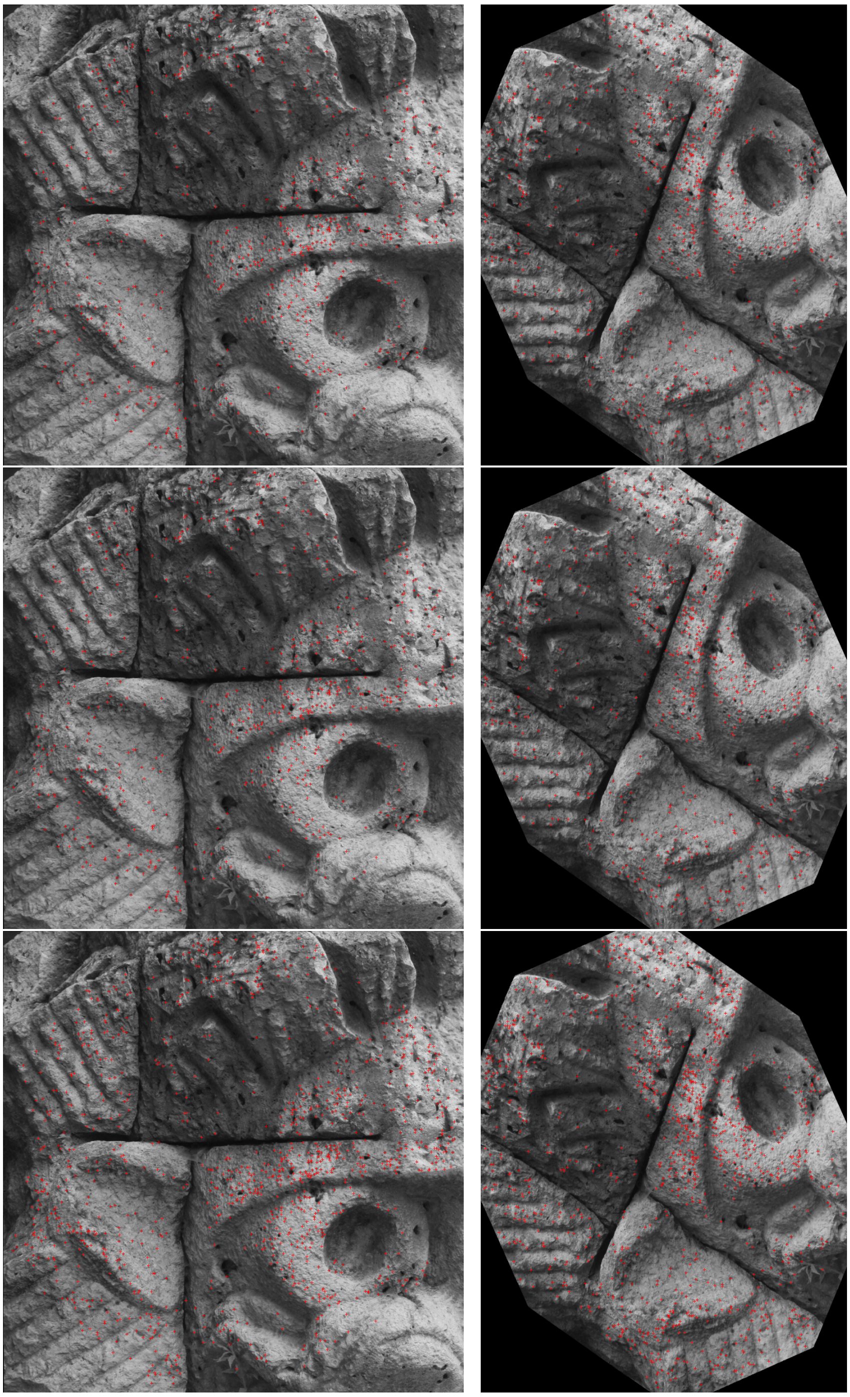

Figure 1: Images plus "inliner" matchings in the first ocatve under affine transformation shown in Eq. (1). First row: Lowe's SIFT method. Second row: improved SIFT by cancelling the sub-sampling in scale space. Third row: iterative SIFT by applying the estimated homography. The left image is the original image and the right image is the original image under the affine transformation in Eq. (1).

		octave -1	octave 0	octave 1	octave 2
translation	$(45, 32)$	1785	909	186	66
	$(45.1, 32.1)$	1334	757	183	65
	$(45.3, 32.3)$	1245	692	181	70
	$(45.5, 32.5)$	1451	691	184	67
	$(45.7, 32.7)$	1233	704	177	64
	$(45.9, 32.9)$	1361	773	179	62
rotation	$15°$	1243	703	190	65
	$25°$	1192	666	173	56
	$35°$	1149	652	165	53
	$45°$	1160	634	158	55
	$55°$	1133	630	158	53
	$65°$	1207	664	161	56
	$75°$	1266	665	191	58
	$85°$	1330	735	201	62
zoom	$2^{1/6}$	872	694	191	61
	$2^{2/6}$	904	757	196	67
	$2^{3/6}$	628	703	177	60
	$2^{4/6}$	432	786	201	65
	$2^{5/6}$	172	724	186	61
	$2^{6/6}$	14	774	202	76
tilt	$2^{1/12}$	1335	752	201	63
	$2^{2/12}$	1123	676	182	56
	$2^{3/12}$	964	606	171	48
	$2^{4/12}$	752	499	135	38
	$2^{5/12}$	505	396	104	28
	$2^{6/12}$	283	227	80	21
affine	$2^{1/12}$	981	561	144	45
	$2^{2/12}$	861	514	132	42
	$2^{3/12}$	756	454	119	36
	$2^{4/12}$	605	370	95	33
	$2^{5/12}$	436	267	82	27
	$2^{6/12}$	236	183	54	19
homography	$2^{1/12}$	916	489	122	34
	$2^{2/12}$	845	482	111	39
	$2^{3/12}$	671	434	110	31
	$2^{4/12}$	527	365	99	29
	$2^{5/12}$	376	249	84	26
	$2^{6/12}$	223	167	56	15

Table 1: The number of “inlier” matchings under different transformations in different octaves for Lowe’s SIFT.

		octave -1	octave 0	octave 1	octave 2
translation	$(45, 32)$	1785	935	229	79
	$(45.1, 32.1)$	1387	807	201	72
	$(45.3, 32.3)$	1263	781	201	69
	$(45.5, 32.5)$	1475	906	224	72
	$(45.7, 32.7)$	1250	786	198	61
	$(45.9, 32.9)$	1376	820	200	63
rotation	$15°$	1257	761	194	53
	$25°$	1199	745	190	62
	$35°$	1180	719	172	57
	$45°$	1166	696	172	61
	$55°$	1150	703	168	59
	$65°$	1229	737	176	63
	$75°$	1274	765	189	62
	$85°$	1344	822	215	66
zoom	$2^{1/6}$	900	759	202	60
	$2^{2/6}$	939	877	217	83
	$2^{3/6}$	652	776	194	64
	$2^{4/6}$	447	823	232	76
	$2^{5/6}$	196	749	180	63
	$2^{6/6}$	22	778	213	75
tilt	$2^{1/12}$	1339	829	209	73
	$2^{2/12}$	1137	749	195	64
	$2^{3/12}$	986	671	177	60
	$2^{4/12}$	773	583	142	48
	$2^{5/12}$	508	456	104	42
	$2^{6/12}$	284	293	73	31
affine	$2^{1/12}$	996	651	156	49
	$2^{2/12}$	865	576	127	42
	$2^{3/12}$	776	511	112	27
	$2^{4/12}$	617	440	105	24
	$2^{5/12}$	438	332	88	18
	$2^{6/12}$	241	209	70	18
homography	$2^{1/12}$	952	529	118	31
	$2^{2/12}$	889	532	118	34
	$2^{3/12}$	710	492	114	32
	$2^{4/12}$	550	419	105	29
	$2^{5/12}$	396	302	85	20
	$2^{6/12}$	236	185	61	15

Table 2: The number of "inlier" matchings under different transformations in different octaves for improved SIFT by canceling the sub-sampling in scale space.

		octave -1	octave 0	octave 1	octave 2
zoom	$2^{1/6}$	1330	677	207	55
	$2^{2/6}$	1435	554	152	38
	$2^{3/6}$	1262	427	126	24
	$2^{4/6}$	1066	339	104	21
	$2^{5/6}$	895	245	88	17
	$2^{6/6}$	761	209	69	13
tilt	$2^{1/12}$	1541	860	232	80
	$2^{2/12}$	1500	790	225	68
	$2^{3/12}$	1345	735	219	65
	$2^{4/12}$	1311	675	211	59
	$2^{5/12}$	1194	641	210	57
	$2^{6/12}$	1066	603	177	46
affine	$2^{1/12}$	1105	646	164	58
	$2^{2/12}$	1090	580	160	51
	$2^{3/12}$	1088	551	147	51
	$2^{4/12}$	1074	536	146	47
	$2^{5/12}$	1018	488	134	44
	$2^{6/12}$	948	453	123	44
homography	$2^{1/12}$	1591	825	217	83
	$2^{2/12}$	1634	843	221	87
	$2^{3/12}$	1811	838	227	80
	$2^{4/12}$	1944	831	226	88
	$2^{5/12}$	1902	783	210	83
	$2^{6/12}$	1933	731	212	83

Table 3: The number of "inlier" matchings under different transformations in different octaves for iterative SIFT by applying the estimated homography.

Page 171: I'm not entirely sure what's done here (section 8.3.3). Is the homography always applied to the image with larger scale or always the one with smaller scale, or arbitrarily? Which residuals are evaluated exactly: residuals in the original image space or in the space of the images warped by the homography?

Page 171, the homography is always applied on the image with larger scale and the residuals is evaluated in the original image space.

I agree with all the minor comments and will modify them in the thesis. For the condition of Theorem 2 and Lemma 1, it also seems to me that 3 non-collinear points should be sufficient. But I have not a proof for that. I will think it over and modify it if I find the proof.