Modelling image quality using estimated contrast, sharpness and noise

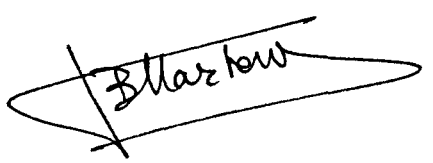
Institute for Perception Research
PO Box 513, 5600 MB Eindhoven

15.08.1996

Rapport no. 1118

Modelling image quality using estimated contrast, sharpness and noise

W.M.C.J. van Overveld



Voor akkoord: Dr.ir. J.B. Martens

Modelling image quality using estimated contrast, sharpness and noise

W.M.C.J. van Overveld

Summary

In this report we describe a simple model for perceived image quality, based on objective estimates of three perceptual attributes: contrast, sharpness and noisiness. The estimates are based on the work of Kayargadde [Kay95].
In this model, image quality is modelled as a polynomial combination of these three estimates. The coefficients are computed from experimental data on images that were varied with respect to (physical) contrast, sharpness and noise.
We investigate the performance of such models by comparing the coefficients found for three different scenes, and by looking at the amount of variance explained by the models. A simple quadratic model with only three terms is shown to perform much better than a linear model, and it is not much worse than a full quadratic model with nine terms.

1. Introduction

It is known that brightness contrast, perceived sharpness and noisiness play an important role in the judgement of image quality for medical images (cf. [Ove94], [Ove95]). Objective measures for these three image attributes have been developed by Kayargadde and others (cf. [Kay95], [KM94], [KMR95], [KM96a], [KM96b], [Ove96]).
We use such estimates to develop and test a very simple model for image quality, to see whether it is possible to predict "average" perceived quality from the three objective measurements. We do not attempt to model individual differences.
Section 2 briefly describes the theory behind three objective measures used: contrast index, blur index and noise index. Section 3 describes an experiment in which we measured quality as a function of three physical parameters, where those physical parameters were expected to correlate well with perceived contrast, sharpness and noise, respectively. We also present the objective measures as computed for each of the stimuli used in that experiment.
Section 4 describes the model: how the measured experimentally assessed quality scores are modelled as a function (either a linear or a quadratic one) of the three objective measures.
Section 5 is devoted to discussion and conclusions.

2. The contrast, blur and noise indices

This section is devoted to a short description of the three indices that are used as estimates of the perceived contrast, unsharpness and noisiness, respectively. More details can be found in [Kay95], [KM94], [KM96a], [KM96a] and KMR[95].
The computation of all three indices starts with a grey-value description of a digital image. This image is assumed to be displayed on a monitor with known grey-value to luminance characteristics, or it is printed on film using a hard copy unit with known grey-value to optical density characteristics. The image is assumed to be viewed in known lighting conditions, and in the case of hard copies, the luminance of the view box is also assumed to be known. All of these assumptions are part of the "display model" (see Figure 1). Using this model, the luminance image can be computed from the grey-value image.
The next step uses a model of the front end of the human visual system. This includes some blur (due to a combination of viewing distance and eye optics), which is modelled as a Gaussian blurring filter with spread b_o. Next, a compressive non-linearity transforms the values in the luminance domain to an internal representation of brightness (some form of the Weber-Fechner law, or Stevens law, as it is called in psychophysical literature; in fact, the CIE lightness model is used, cf. [Hun78]). Finally, some internal noise is added with spread n_o. At this stage (the output of the front-end model) we have arrived at a "brightness" image. The three different estimates of contrast, unsharpness and noisiness are obtained on this representation of the image, as shown in Figure 1.

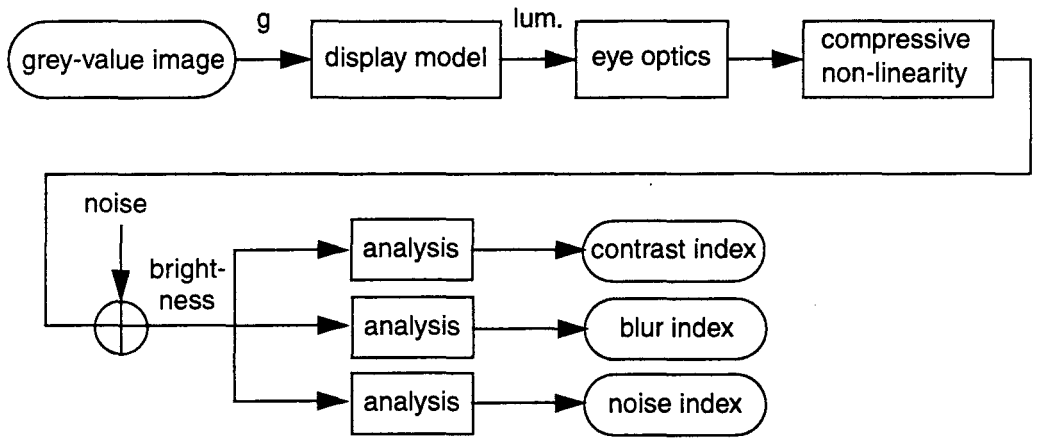


FIGURE 1. Computation of contrast, blur and noise indices.

The contrast index [KMR95] is based on the histogram of the brightness values. The (global) contrast index is defined as the inverse of the slope of the cumulative brightness histogram. In the case of a uniform brightness histogram, this index is proportional to the width of the brightness histogram, which is intuitively related to the perceived contrast. The slope of the cumulative histogram is computed using linear regression of the part of the histogram between 5% and 95%. In [Ove96], we have extended this notion to a "local contrast index", but since this did not always give satisfactory results, we only use the global contrast index in this study.

The blur index ([KM94], [KM96b]) is computed using Gaussian derivative operators. An image is analysed within Gaussian windows of a given spread σ_a. Within such a window, local (Gaussian) derivatives up to order 3 are computed. Using these derivatives, it is possible to estimate the five parameters that characterize an ideal edge blurred by a Gaussian filter: the location of the edge relative to the analysis window, its orientation, mean value and height, and finally the spread of the blurring filter, σ_b. σ_b is an estimate for the blur of the given edge.
The blur index is computed from the σ_b estimates of various edges in the image. In fact, only those estimates are used which can be considered as reliable: first of all, we must be sure that there is indeed an edge at the given location (by thresholding the amount of one-dimensional energy relative to the total energy), and secondly, the distance of the edge to the center of the analysis window and σ_b should be comparable in size to σ_a which can be ensured by setting another threshold.

Now a global estimate for the image blur, say $\bar{b}$, can be obtained from all "reliable" σ_b estimates, by minimizing the weighted error between the global estimate and the local ones. The weighting takes care of the confidence in the local estimate and is based on the local edge gradient. The estimated "global blur" $\bar{b}$ is still an estimate of a physical parameter (despite the fact that it is computed from a brightness image). This can be transformed to an estimate of perceived blur through

$$I_b = 1 - \sqrt{b_o / \bar{b}},$$

where b_o is the spread of the intrinsic blur in the optical system (note that this is included in the estimate of $\bar{b}$ because the brightness image used as input to the estimation algorithm already incorporates optical blur). I_b is called the blur index.

The noise index [KM96a] is also based on local derivative analysis. In this case, the image is analysed using Gaussian windows of two different spreads, say σ_1 and σ_2. The first-order derivatives in both directions are computed, and the amplitude

$$|f_1| = \sqrt{f_{0,1}^{\ 2} + f_{1,0}^{\ 2}}$$

is used to determine the noise variance. Under some assumptions (among others, that the image contains enough locally uniform regions), the peak of the distribution of $|f_1|$ occurs at location β, which is the spread of the first-order derivative coefficients in uniform, noise-only regions. Assuming that the noise is white with standard deviation σ_n and this noise is blurred by a Gaussian filter of spread $\sigma = l_c$ it can be shown that

$$\beta_i = \frac{\sigma_n}{2\sqrt{\pi}} \cdot \frac{\sigma_i}{l_c^{\ 2} + \sigma_i^{\ 2}},$$

in which the subscript i takes the value 1 or 2, denoting the two different analysis windows. Thus we have two equations, allowing us to estimate both σ_n and l_c. Kayargadde and Martens have found that σ_n correlates well with the perceived noisiness, so that the noise index is based on this estimate. The noise index, I_n, is computed from the "physical" estimate of noise in the same way as the blur index:

$$I_n = 1 - \sqrt{n_o / \bar{n}},$$

in which n_o is the spread of the intrinsic noise in the optical system and $\bar{n}$ is the estimated noise spread including the intrinsic noise. In fact, we used a slightly different version of this formula: we did not actually include the intrinsic noise in the front-end model because then we would have needed many runs of the algorithm (with different realisations of the intrinsic noise) to estimate σ_n. Instead, we assumed $n_o = 0$ in the front-end model, computed σ_n as an estimate of the noise present in the image (excluding the internal noise) and afterwards assumed that the overall standard deviation of the internal plus the external noise could be written as

$$\bar{n} = \sqrt{n_o^{\ 2} + \sigma_n^{\ 2}}.$$

3. Experiment: quality scaling and objective measures

In this chapter, we briefly recapitulate an experiment already described in [Ove95]. In this experiment, three physical parameters were varied: contrast (grey value range), sharpness (width of a Gaussian blurring kernel) and noise (simulated X-ray dose). These variations were applied to three angiographic X-ray images, named cer, kid and leg. The images will be referred to as "scenes". Four different levels were chosen for each of the physical parameters.
We first varied the amount of noise using a computer simulation of a varying X-ray dose. The noise levels are indicated by 1 to 4 where 1 is the lowest noise level (the highest X-ray dose). Next the contrast variation was applied, consisting of a linear scaling of the grey value range with varying window width, ensuring that the average grey value was kept constant over the variation. The widths (referred to as contrast levels) were 255 (maximum contrast), 200, 150, and 100. Finally, the images were blurred with a Gaussian filter with a spread of 0 (no blur), 1, 2 or 3 pixels standard deviation. These will be called blur level 0 to 3, respectively.

Quality scaling experiment

All images (64 versions per scene) were viewed in a dimly lit room, on a view box having a luminance of L_{box} = 2500 cd/m^2. The viewing distance was 50 cm. Every subject viewed the different versions of a scene in randomized groups of 16 images at a time. The three scenes were not directly compared with each other. The subjects had to express the quality of each image on a scale from 1 to 10: both in terms of appreciation and "diagnostic value", i.e. the ability to see details. We used 8 non-experts as subjects. We changed the quality scores into z-scores per subject (with zero average and unit standard deviation) and averaged these over the subjects. The average scores were transformed back to the original scale.
The results can be found in Figures 2-4, which depict the same results as Figure 12 of [Ove95].

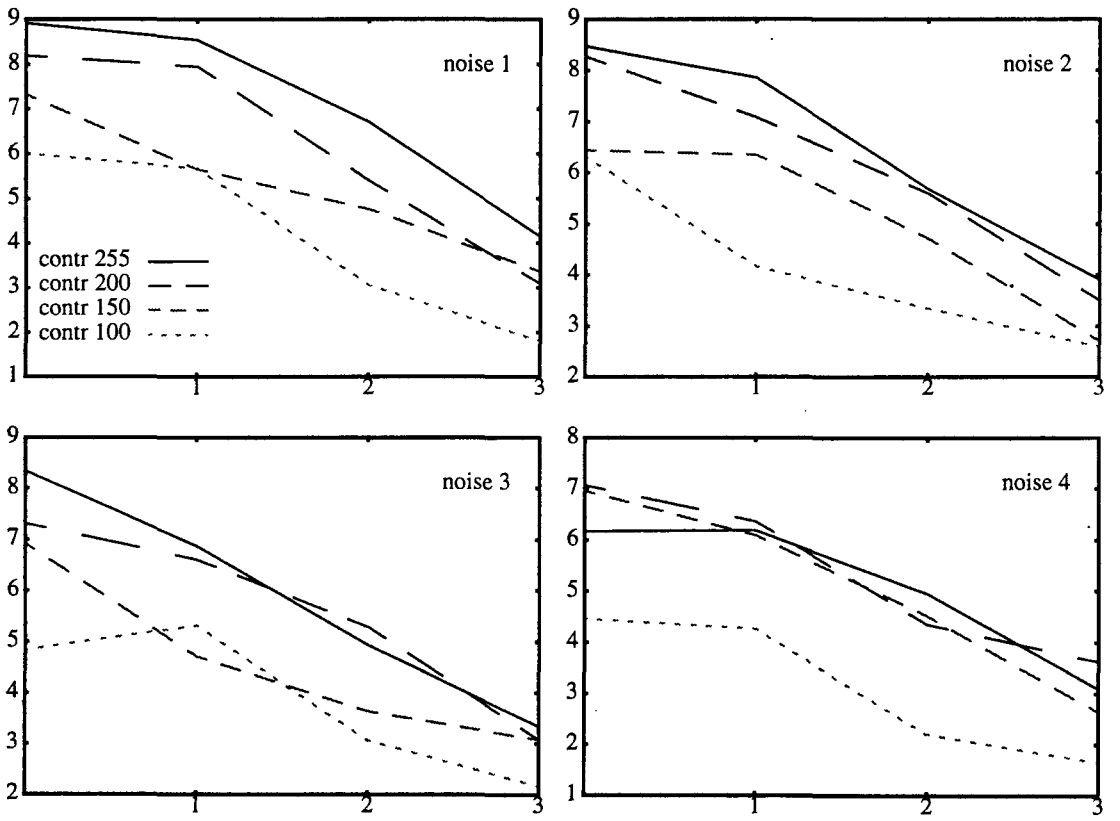


FIGURE 2. Quality scores for scene "cer", as a function of blur level (increasing in four steps, from left to right on the horizontal axis). The four different plots are for the different noise levels, and different curves in one plot are for different contrast levels. The size of the dashes indicates the contrast level: the shorter the dashes, the lower the contrast.

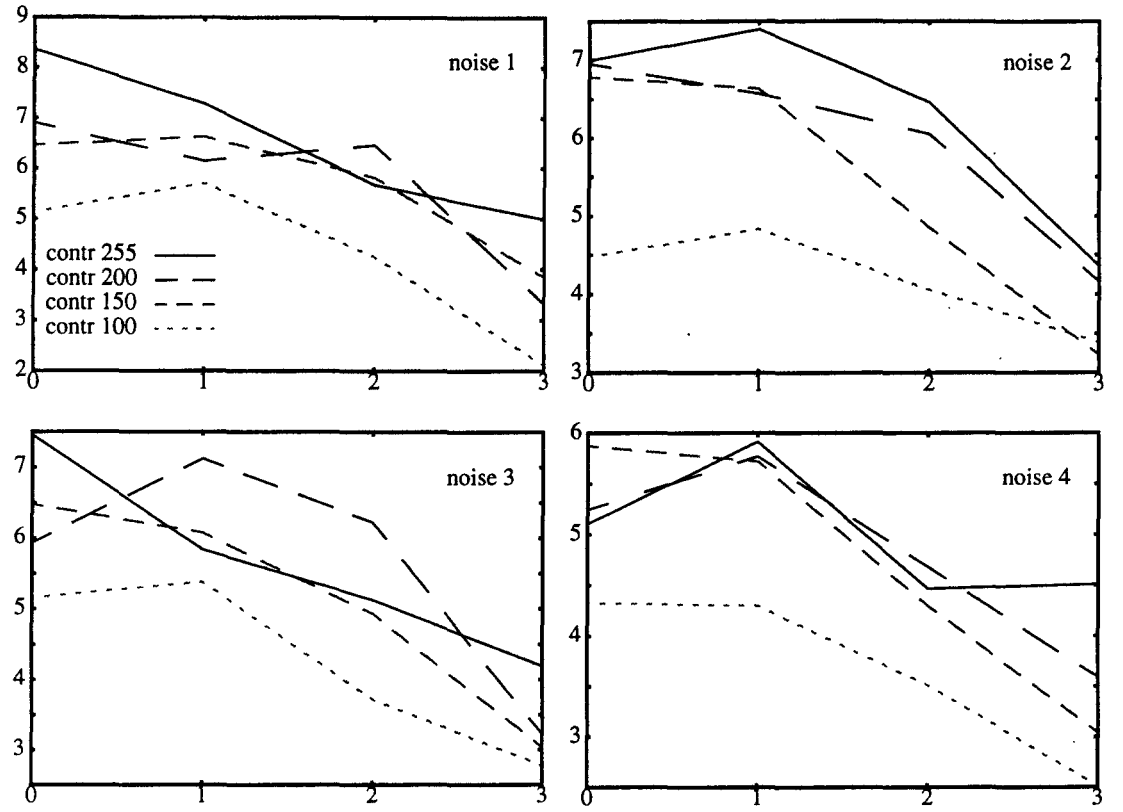


FIGURE 3. Quality scores for scene "kid", as a function of noise level. See the caption of Figure 2 for further description.

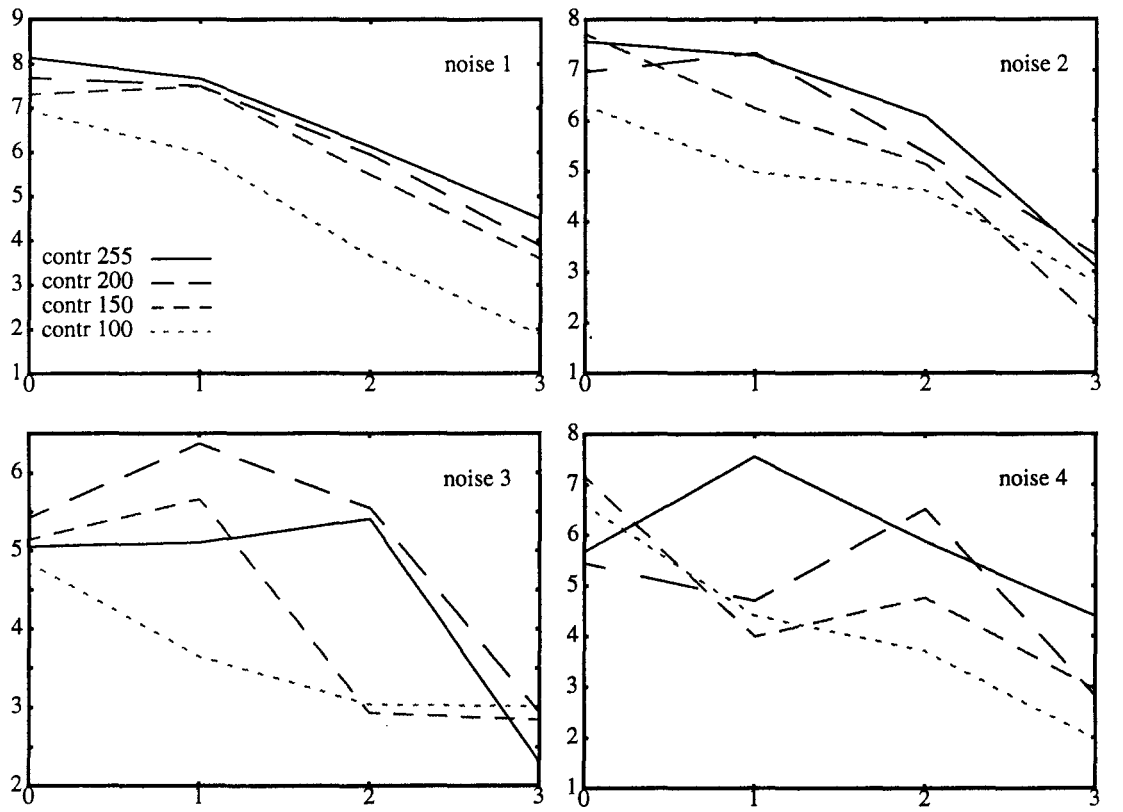


FIGURE 4. Quality scores for scene "leg", as a function of noise level. See the caption of Figure 2 for further description.

Objective measures

Figures 5-7 show the contrast indices, computed for each of the images used in the experiment. As in [Ove96], the black borders of the images were excluded from the histogram before the computation.

The estimates of the blurring spread, $\bar{b}$, are shown in Figures 8-10. These were computed as described in Section 2, with the parameter values chosen as in [KM96b], except for the fact that we used a Gaussian analysis window with $\sigma_a = 2$ instead of a binomial window. To avoid problems with the black borders of the image (which could be regarded as ideally sharp edges, thus erroneously lowering the overall estimate of the blurring spread) we only used subimages of the original images. The square subimages were chosen as large as possible without including parts of the black border, and they contained sufficiently many representative parts of "real" edges in the image. To convert the blurring spread to the blur index, we used the value of b_0 that was used in the front-end model, namely 0.8067 pixels (this corresponds to 0.65 minarc as used in [KM96b], given the viewing distance and size of our images).

Estimated noise spread σ_n is shown in Figures 11-13. For this, we used the same subimages as used for the blur estimates (note that the black borders would count as completely noise-free regions). Again we used Gaussian analysis windows instead of binomial ones, as stated in [KM96a]. We used five scales of analysis windows: 1, 1.5, 2, 3 and 4, and estimated σ_n from appropriate pairs of these scales as described in section 2.2 of [KM96a].
In the computation of the noise indices, the value of n_0 was chosen to be 2 brightness units. This is in the same order of magnitude as the values used in [KM96a], and it is also approximately equal to the smallest value of σ_n found for our images. Thus we assume that the noise in the images with the highest X-ray dose is at the same level as the internal noise.

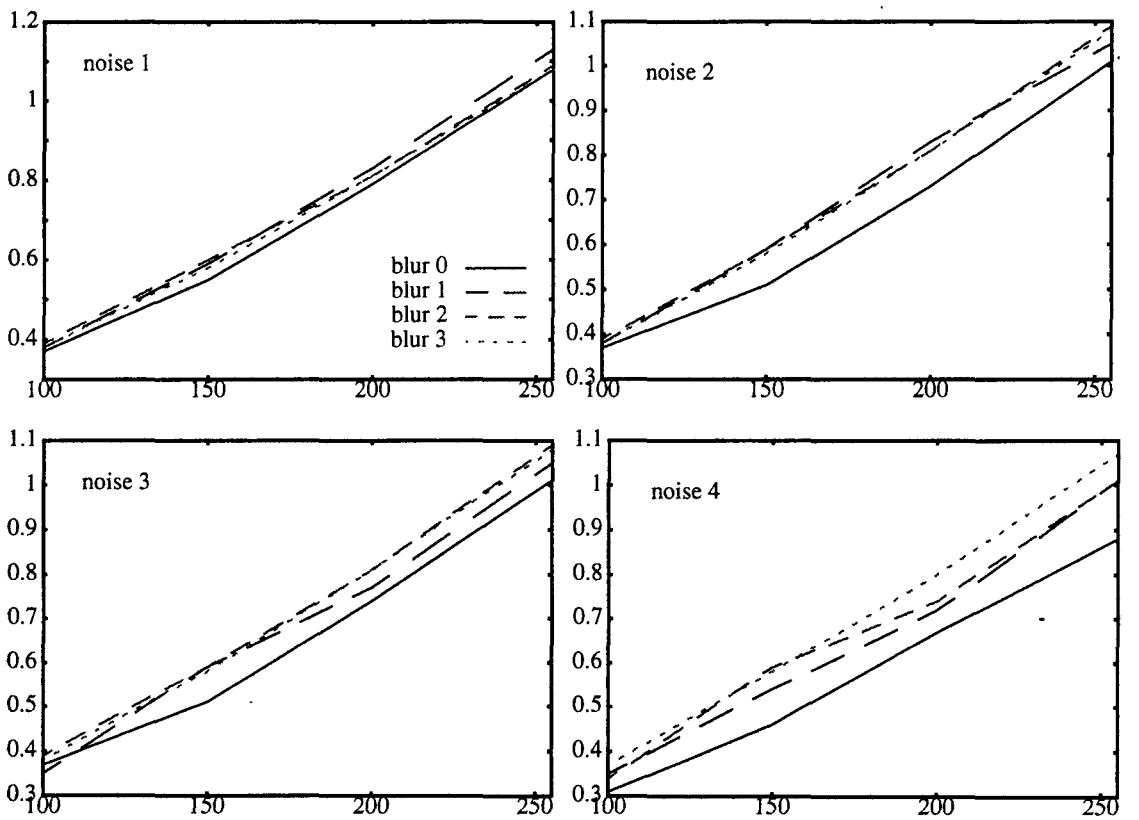


FIGURE 5. Contrast indices for scene "cer", as a function of contrast level. The four different plots are for the different noise levels, and different curves in one plot are for different blur levels. The size of the dashes indicates the blur level: the shorter the dashes, the more blur.

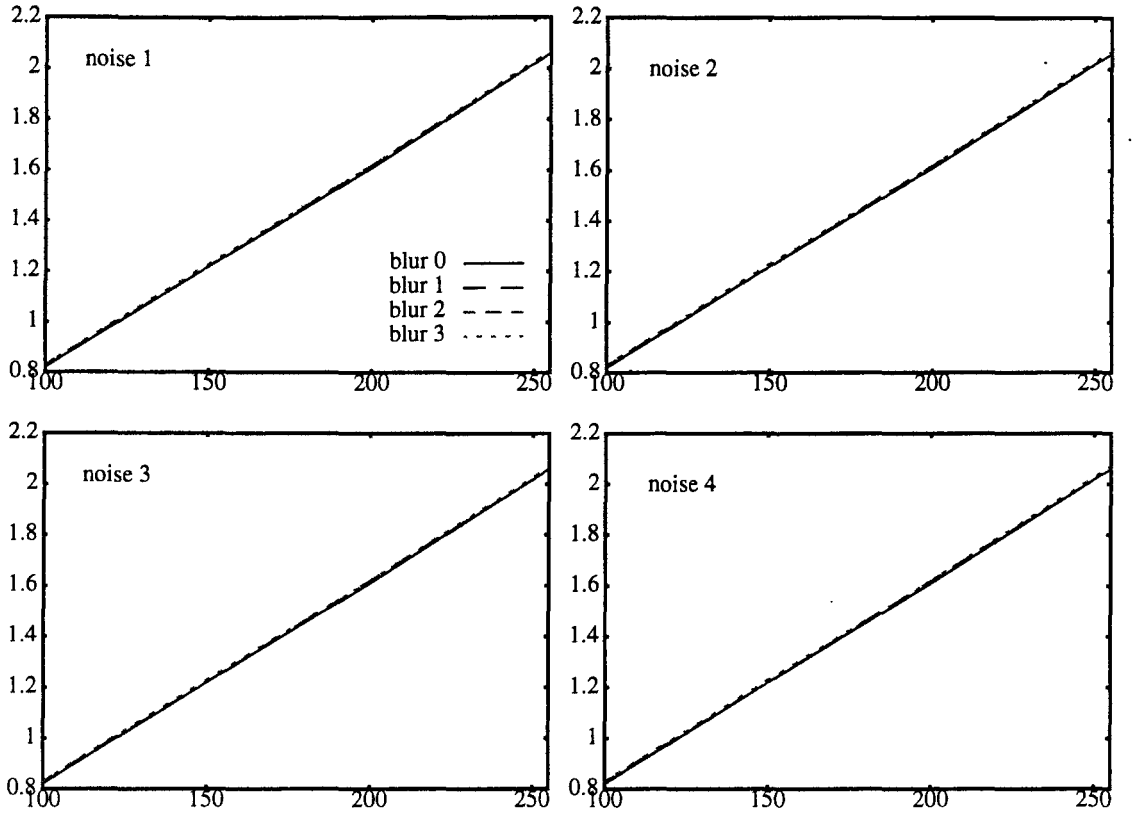


FIGURE 6. Contrast indices for scene "kid", as a function of contrast level. See the caption of Figure 5 for further description.

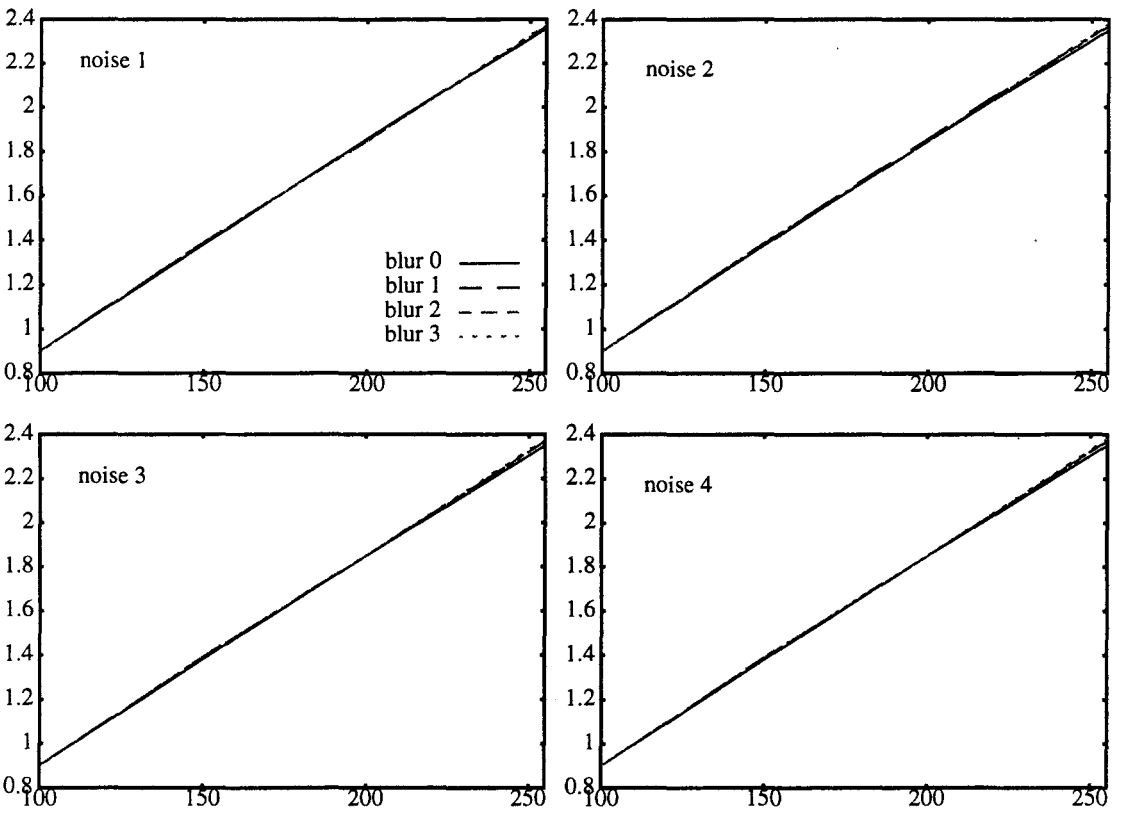


FIGURE 7. Contrast indices for scene "leg", as a function of contrast level. See the caption of Figure 5 for further description.

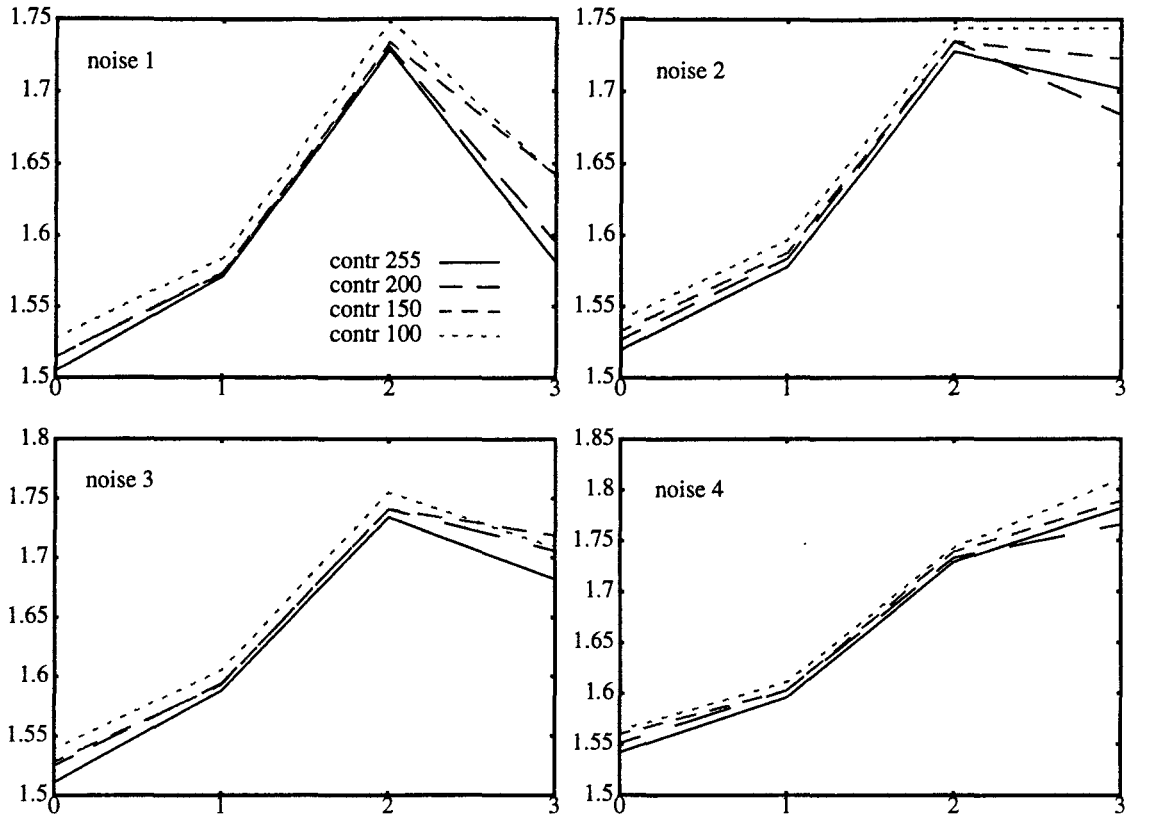


FIGURE 8. Estimated blur spread for scene "cer", as a function of blur level. The four different plots are for the different noise levels, and different curves in one plot are for different contrast levels. The size of the dashes indicates the contrast level.

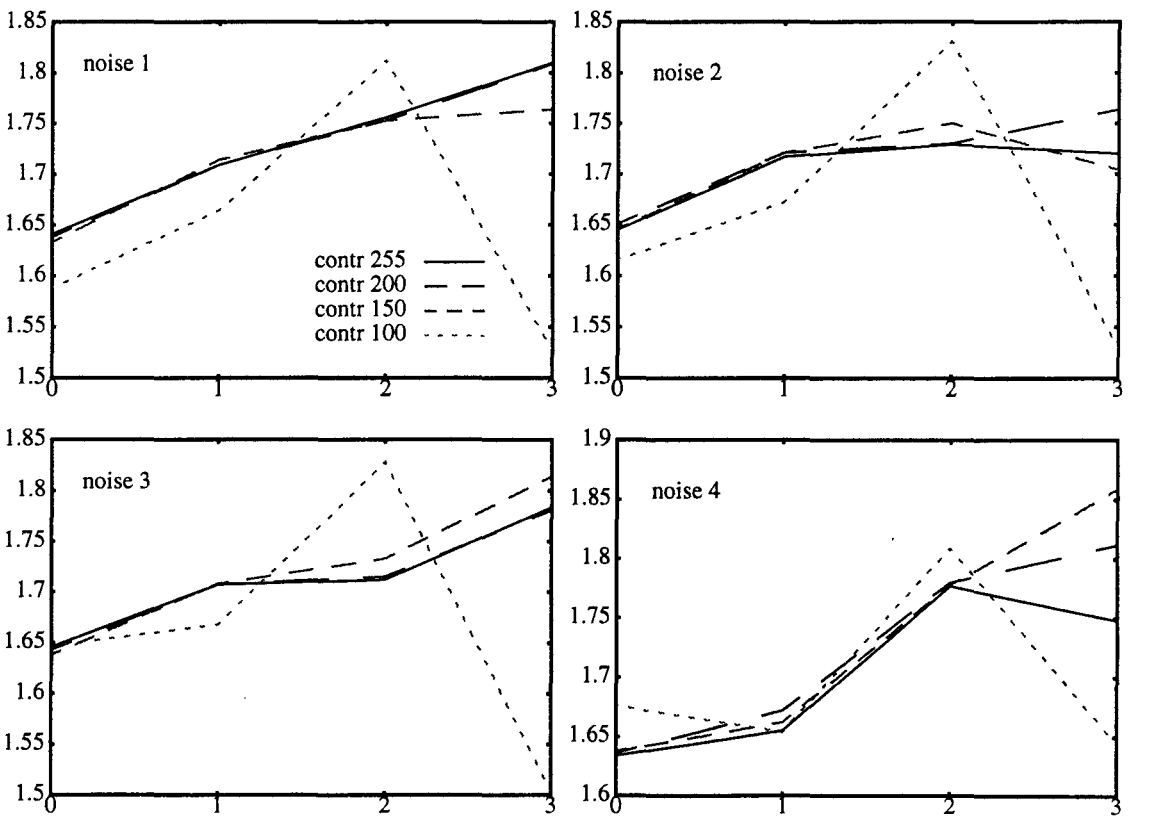


FIGURE 9. Estimated blur spread for scene "kid", as a function of blur level. See the caption of Figure 8 for further description.

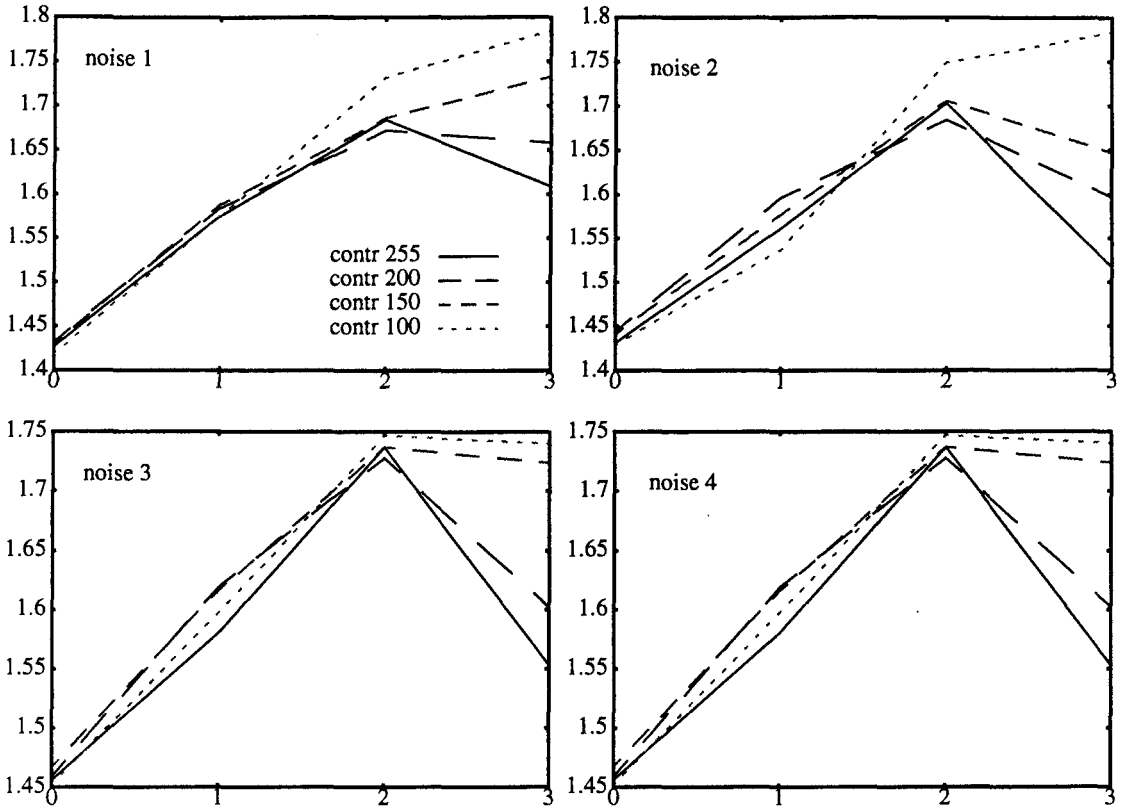


FIGURE 10. Estimated blur spread for scene "leg", as a function of blur level. See the caption of Figure 8 for further description.

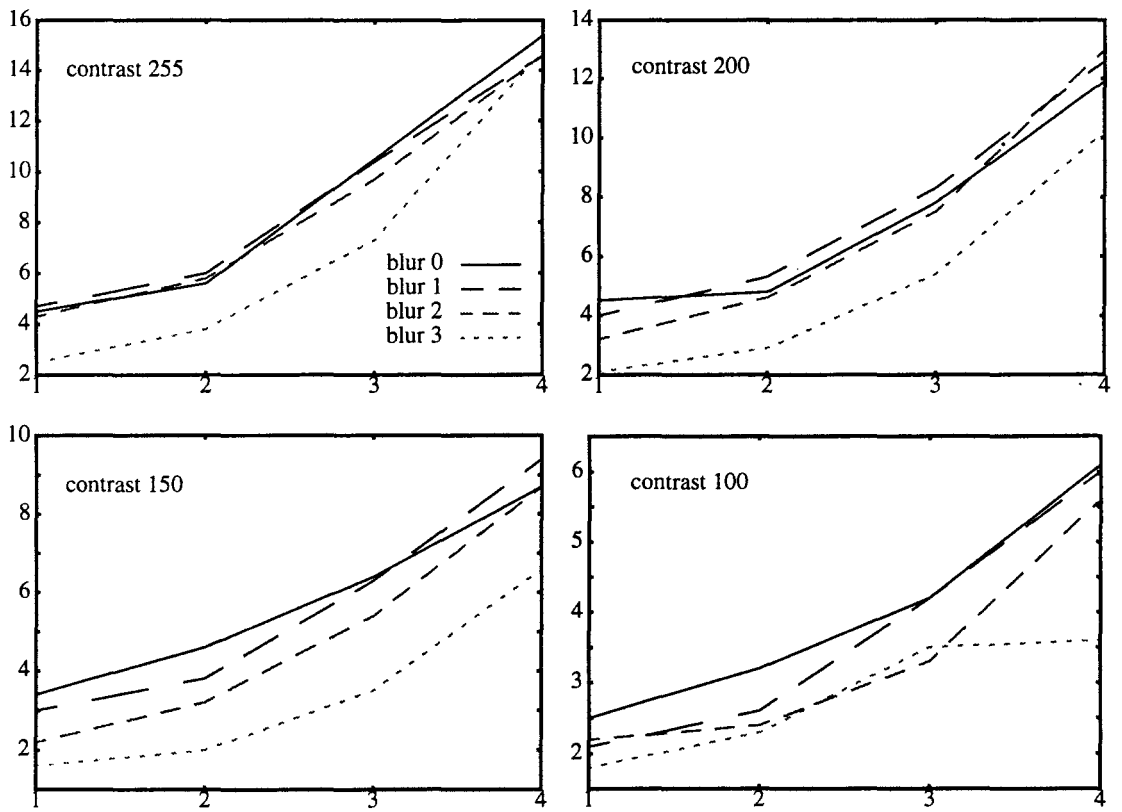


FIGURE 11. Estimated noise spread for scene "cer", as a function of noise level. The four different plots are for the different contrast levels, and different curves in one plot are for different blur levels. The size of the dashes indicates the blur level.

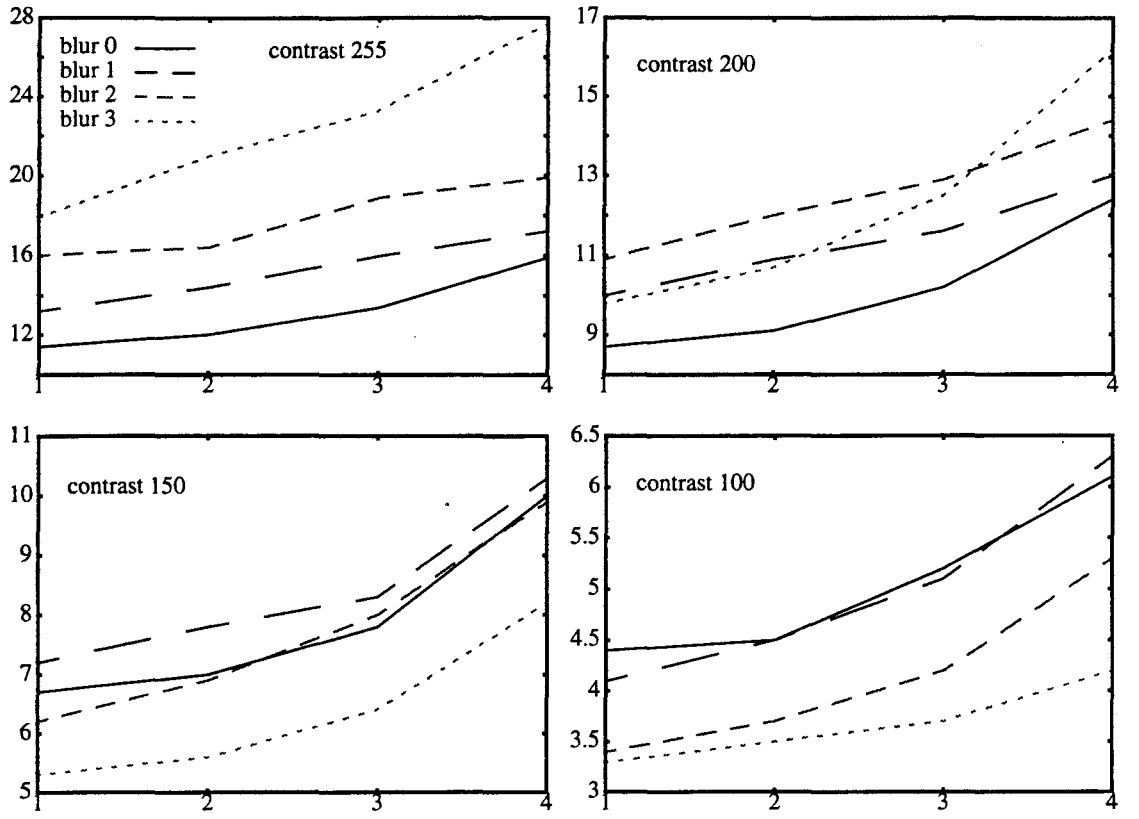


FIGURE 12. Estimated noise spread for scene "kid", as a function of noise level. See the caption of Figure 11 for further description.

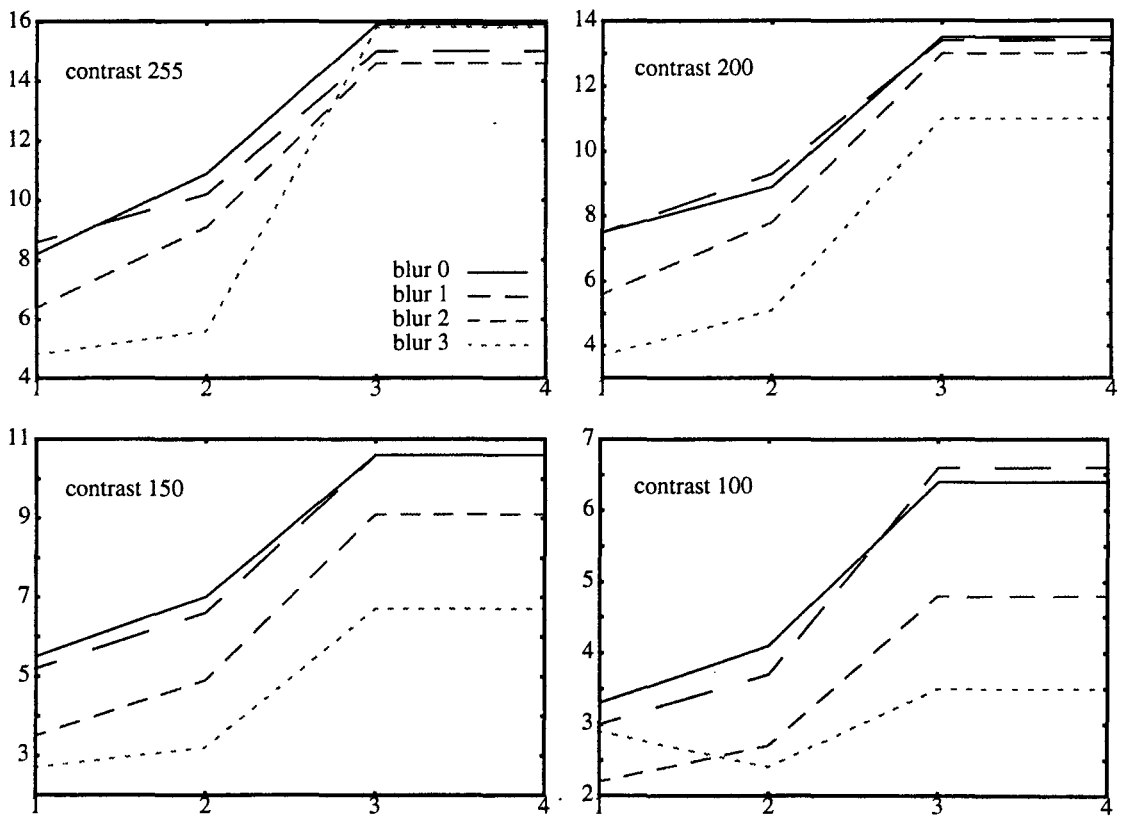


FIGURE 13. Estimated noise spread for scene "leg", as a function of noise level. See the caption of Figure 11 for further description.

Discussion

The objective measures as computed for the 3 x 64 images used in the experiment give rise to the following observations.

As already observed in [Ove96], the effect of the varying window width is reflected very well in the contrast index. We see that the estimate is insensitive to noise and blur, although the contrast index drops a little when the subtraction image "cer" is extremely noisy.

The estimated blurring spread describes the varying blur level well when the blurring kernel is not too large. For blur levels 0 to 2, we see that the estimated blur indeed increases with the blur level. For these levels, we also see that the estimates are quite insensitive to noise variations and contrast variations (except for contrast level 100 in the "kid" scene).
For blur level 3 (the right-most points in Figures 8-10), the estimated blur behaves much less predictable. In the case of scene "cer", we see a noise effect and for the other two scenes, we see an effect of the contrast level. The fact that the estimates don't agree with the blur level can be explained from the selected size of the analysis window. We used a window with a spread of 2, while the highest blur level corresponds to a blurring kernel with spread 3. To estimate such an amount of blur correctly, a larger analysis window would be needed. If we used a larger window, the smaller amounts of blur would be estimated less accurately and the algorithm would be computationally more demanding. However, since blur level 3 introduces an amount of blur much larger than that actually encountered in X-ray images, we decided to stick to an analysis window with a spread of 2. We will just neglect the blur estimations for blur level 3 in the image quality model presented in the next section.

The noise spread estimates, finally, are seen to increase with the noise level. The estimates also increase with the contrast level. Increasing blur has a tendency to lower the noise spread estimate, except for the "kid" scene at high contrast levels where the effect is just the opposite. It is not clear why the "kid" scene behaves differently from the other two in this respect.

4. Quality model

This chapter explains how the objective measures are used to find a quality model. The model prediction of quality will be a polynomial function of the contrast, blur and noise indices. The simplest case is the linear model:

$$q = a_0 + a_c \cdot I_c + a_b \cdot I_b + a_n \cdot I_n.$$

In this formula, I_c, I_b and I_n are the contrast, blur and noise index, respectively; a_c, a_b and a_n are the coefficients with which each of these indices has to be multiplied. a_0 is an additive constant. Using linear regression, the best fitting constants a_0, a_c, a_b and a_n and can be determined.
The goodness-of-fit may be expressed in terms of the squared correlation coefficient, r^2: this indicates which percentage of the variance in q can be explained from the model prediction.
We still have a few options for the construction of such a model: we can either use the data of all three scenes in one single regression, or we can make three separate models for each of the scenes. It seems to make more sense to make three separate fittings, because we cannot say anything about the comparison of quality scores across scenes: scenes were never compared to each other in the experiment. However, we can deal with this by assuming that the differences in quality scores between scenes can be modelled by a linear scaling: $q' = a_s + b_s q$ where a_s and b_s are scene dependent. Thus the parameters a_s and b_s (for three values of scene s) enter in the regression formula. Here $a_s = 0$ and $b_s = 1$ can be fixed for one of the scenes (in our case, we fixed these for scene "cer").
The second choice we have to make is whether to use the data from all blur levels, or whether to exclude the data points corresponding to blur level 3. As discussed in the previous section, we cannot expect to find a good fit when we include blur level 3, because the blur index does not even correlate well with the spread of blurring kernel in this case. We can expect to find a better correlation if we exclude blur level 3, if only because we have to fit the model to fewer points.
In Table 1 below we show the results for the four cases: data pooled over the scenes or separate data sets for each of the scenes, and with or without blur level 3.

Table 1: coefficients in the linear quality models

	with blur level 3				**without blur level 3**			
	cer	**kid**	**leg**	**pooled**	**cer**	**kid**	**leg**	**pooled**
a_0	24.0972	10.7763	13.6695	24.0301	22.1935	15.7929	13.2309	23.5889
a_c	2.35180	1.06689	0.60794	2.24429	3.97134	2.19717	1.74935	3.71460
a_b	-71.313	-26.107	-33.868	-71.179	-62.175	-35.346	-28.177	-65.335
a_n	1.86771	1.95261	0.57979	2.13243	-1.8426	-4.2357	-4.4845	-2.5619
r^2	0.70627	0.26953	0.34632	0.47279	0.86494	0.49635	0.52820	0.65201

Even the best of the linear models is not very good in predicting the quality scores. Looking at the "cer" scene when blur level 3 is excluded, we find $r^2 = 0.86494$, but for the other scenes, only about 50% of the variance can be explained. The obvious step to improve the model is to look at a quadratic model. If we allow all possible coefficients to be non-zero, we arrive at the so-called "full" quadratic model:

$$q = a_0 + a_c I_c + a_b I_b + a_n I_n + a_{cc} I_c^2 + a_{bb} I_b^2 + a_{nn} I_n^2 + a_{cb} I_c I_b + a_{cn} I_c I_n + a_{bn} I_b I_n.$$

The fact that we might need to include cross-terms in the model could be expected if we realize that the three indices are correlated. For instance, both the contrast index and the noise index increase with the physical contrast level.The reason for including the squared indices along with the indices themselves is less obvious. It is inspired by the work of De Ridder [Rid92] who

showed that different types of perceived image impairments (such as blur or noisiness) are combined into an overall quality impression (q) using a euclidean summation rule. In that case we should be able to model q^2 as a sum of squared contrast, blur and noise indices. For the moment we stick to the model for q as such; we come back to a model for q^2 later in this section.
Again we consider four different model fits: for separate scenes or pooled over scenes, and with or without blur level 3. Table 2 shows the results.

Table 2: coefficients in the full quadratic models

	with blur level 3				**without blur level 3**			
	cer	**kid**	**leg**	**pooled**	**cer**	**kid**	**leg**	**pooled**
a_0	81.3181	-141.90	50.2875	24.9833	42.4062	-177.27	4.85345	24.2939
a_c	0.23829	-2.1688	-7.9976	-9.1540	8.94127	-0.2646	4.38504	1.98576
a_b	-495.88	876.517	-272.56	-133.79	-211.71	1167.83	55.1083	-80.258
a_n	37.5595	72.9531	20.5859	59.0037	0.28251	23.5548	-22.392	1.37888
a_{cc}	-5.0332	1.53970	-2.0671	-1.0847	-2.5841	2.46847	-1.7152	-0.4334
a_{bb}	738.899	-1371.1	340.851	157.253	237.983	-1939.3	-235.94	7.63346
a_{nn}	-23.611	-9.8927	-48.470	-30.500	7.39162	39.0982	-18.081	-1.6105
a_{cb}	22.5221	17.5814	31.5306	40.2427	12.7440	52.6007	4.75162	21.1145
a_{cn}	7.63424	-10.523	14.5809	4.41696	-12.149	-35.436	4.09201	-7.9514
a_{bn}	-78.016	-169.65	-5.1550	-123.31	-1.1384	-100.09	89.3852	10.2159
r^2	0.78003	0.64312	0.47863	0.53453	0.91368	0.81647	0.69253	0.76229

Obviously, all fits are better than those in the corresponding linear models. We also see that the variation of coefficients across scenes is much higher than for the linear models. This is due to the fact that with this many degrees of freedom, one is likely to find nearly optimum solutions in various regions of the parameter space. It follows that such a complex model is very hard to interpret.
To find a compromise between the good fit of the full quadratic model and the simplicity of the linear model, we now investigate quadratic models in which we allow only three non-zero coefficients plus an intercept (a_0). Thus we have the same number of coefficients as in the linear model, but we can select them from a larger set. We call this a "reduced quadratic model".
Since the "separate scenes, without blur level 3" model outperforms the other ones (at least for the quadratic models), we concentrate on this case. When all possible reduced models are tested for this case - there are 84 of them - it turns out that the model given in Table 3 gives the best "total fit". By "total fit", we mean the sum of the r^2 terms for the three scenes. In fact, the model of Table 3 also gives the highest r^2 for each separate scene, except for one other model which gave a marginally higher r^2 for the "cer" scene (but significantly lower r^2's for the "kid" and "leg" scenes).
Table 3 also gives coefficients for the other three cases, i.e. including blur level 3, and pooling over scenes. These are the results of fitting the same reduced model to the data points of those three cases.

Table 3: coefficients in the reduced quadratic models

	with blur level 3				without blur level 3			
	cer	**kid**	**leg**	**pooled**	**cer**	**kid**	**leg**	**pooled**
a_0	14.7389	7.93676	9.65860	15.1963	14.1292	10.8072	10.1387	14.6749
a_{bb}	-126.17	-63.007	-69.141	-130.58	-126.85	-118.10	-90.804	-131.54
a_{cb}	5.27855	12.8526	3.35737	6.33213	19.2050	35.7783	14.4832	22.0934
a_{cn}	2.18507	-2.9805	-0.3285	1.18104	-3.7472	-12.026	-4.2163	-6.1986
r^2	0.69508	0.27240	0.35171	0.46340	0.88105	0.71253	0.62302	0.72873

For the case of separate scenes without blur level 3, we find a fairly good fit. We observe that the r^2 value for scene "cer" is approximately halfway between the values for the linear model and for the full quadratic model; for the other two scenes, the r^2 value is much closer to the value of the full quadratic model.
It can also be seen that the coefficients for the three different scenes are much closer to each other than they were in the full quadratic model. Even after pooling over the scenes, the values of the coefficients do not change too much. This indicates that a reduced model with fixed values for the coefficients can be used to predict the quality scores for different scenes. When we use the following values (obtained by averaging the values for the three separate scenes):

a_0 = 11.6917
a_{bb} = -111.918
a_{cb} = 23.1555
a_{cn} = -6.66317,

the model explains 83.8%, 63.6% and 57.9% of the variance for the scene "cer", "kid" and "leg", respectively.
The terms in the model can be interpreted as follows. The fact that the negative coefficient a_{bb} occurs shows the importance of blur in the subjects' quality scores. This was already noted in [Ove95]. The positive value of a_{cb} indicates the positive effect of contrast. When an image is blurred, high contrast appears to be even more important. Interestingly enough, this cross-term plays a more important role in predicting the quality than the contrast index *per se* [1]. The negative coefficient a_{cn} means that noisiness has an especially detrimental effect on the quality when the contrast level is high. Again, this is more important than the noise index alone (even though the noise index itself is already increasing with contrast level).

Finally we investigate a model for q^2, which was our incentive to add the square terms in our model. We try the reduced model given in Table 3 (i.e. a quadratic model in which only a_0, a_{bb}, a_{cb} and a_{cn} may be non-zero), to model q^2. When we only consider the case of separate scenes in which blur level 3 is excluded, we find r^2 values of 0.85899, 0.69957 and 0.59640, respectively. Comparing these to the values in the model for q (0.88105, 0.71253 and 0.62302), it is seen that this reduced model for q works slightly better than the same model for q^2. One might argue that for q^2, the optimum reduced model could have different non-zero coefficients, but a quick search among the most promising of the 84 possible models did not produce any better models.

1. Note that, even if we were modelling perceived sharpness only instead of quality, we would always need a cross-term for the interaction of contrast and blur, because of the known effect of contrast on perceived sharpness.

5. Conclusions and discussion

A number of conclusions can be drawn from the work described in this report.

We have seen that our objective measures for contrast, blur and noise are able to describe the effects of variation of physical contrast, blur and noise. However, we have also found some interactions: especially the estimate for noise seems to be sensitive to variations in contrast and blur. It would be interesting to try and come up with an improved noise index which is a more robust estimate of the noise standard deviation.
We have not checked whether the blur and noise indices are also able to describe the perceived sharpness and noisiness. For the contrast index, the relation with perceived contrast was investigated in [Ove96]. In the present study, we have only investigated the relation between the estimates and the perceived quality without looking at the intermediate step of perceived image attributes.
As already discussed in Section 3, the blur index does not work well when the blurring spread is too large. When large amounts of blur are expected in an image, one could apply a larger analysis window, but this will not be necessary for everyday purposes.

The regression models presented in Section 4 are able to predict the quality scores to a fair extent (accounting for 60% to 80% of the variance). This holds for models that are separately fit to the data of individual scenes, but also when a single fit is used for all scenes (allowing for a linear scaling of the scores for different scenes). We do however need to discard the data for images with blur level 3, because the blur index is extremely unreliable at this level.
We have shown that a relatively simple model with three second-order terms plus an offset works much better than a linear model with the same number of variables. The terms and the signs of their coefficients could be interpreted in a meaningful way. A quadratic model with 10 parameters was shown to work still better, but this cannot be interpreted any more.

A final word of warning: quality models as presented in this report are not able to describe individual preferences of different subjects. In order to do this, we would need to fit models to the data of separate subjects. This would require an enormous amount of data, because every combination of subjects and scenes would have to be modelled independently. Thus the models presented here are only able to convey an average impression of quality.

References

[Hun78] R.W.G. Hunt, "Colour Terminology". *Color Research and Application 3*, pp. 79-87, 1978.

[Kay95] V. Kayargadde, *Feature extraction for image quality prediction.* Ph.D. Thesis, Eindhoven University of Technology, 1995.

[KM94] V. Kayargadde and J.B. Martens, "Estimation of edge parameters and image blur using polynomial transforms". CVGIP: Graphical Models and Image Processing, vol. 56, pp. 442-461,1994.

[KMR95] V. Kayargadde, J.B. Martens and J.A.J. Roufs, "An objective measure for global brightness contrast". Submitted to Journal of the SID / IPO Manuscript 1113, 1995.

[KM96a] V. Kayargadde and J.B. Martens, "An objective measure for perceived noise". Signal Processing, vol. 49, pp. 187-206, 1996.

[KM96b] V. Kayargadde and J.B. Martens, "Estimation of perceived image blur using edge features", Int. J. of Imaging Systems and Technology, vol.7, pp 102-109, 1996.

[Ove94] W.M.C.J. van Overveld, "The effect of gamma on subjective quality and contrast of X-ray images". Proceedings of SPIE Medical Imaging '94, SPIE vol. 2166: Image Perception. pp. 96-104, 1994.

[Ove95] W.M.C.J. van Overveld, "Contrast, noise, and blur affect performance and appreciation of digital radiographs". Journal of Digital Imaging, vol. 8, no.4, pp. 168-179, 1995.

[Ove96] W.M.C.J. van Overveld, "A local measure for perceived contrast" IPO report no. 1115, 1996.

[Rid92] H. de Ridder, "Minkowski-metrics as a combination rule for digital-image-coding impairments", in: B. E. Rogowitz (Ed): Human Vision, Visual Processing, and Digital Display III: Proceedings SPIE - The International Society for Optical Engineering, San José, vol. 1666, pp. 16-26, 1992.