Proc. IEEE Int. Conf. on Image Processing, Lausanne, Switzerland, September 16-19, 1996, vol. I, pp. 265-268.

A PYRAMID APPROACH TO SUB-PIXEL IMAGE FUSION BASED ON MUTUAL INFORMATION

Philippe Thévenaz and Michael Unser
National Institutes of Health
Bethesda MD 20892–5766, USA

ABSTRACT

We investigate aspects of multi-modal image registration based on a new criterion named mutual information (or sometimes Shannon information). This criterion is intensity-based and requires no landmarks; hence, its application can be automated without resorting to segmentation. We present a form amenable to derivation with respect to the geometric transformation parameters (here: affine transformation). This form involves Parzen windows; we explore the dependence of the registration accuracy on these windows and propose that they be tuned to each resolution level in a pyramid approach. We conduct experiments and show that both the window width and the number of windows is relevant. In addition, we show that it is beneficial to use a spline-based high-order interpolation scheme for applying the geometric transformation.

1. INTRODUCTION

In a previous paper [7], we were able to show that a sub-pixel registration method based on a coarse-to-fine multi-resolution approach could achieve the double goal of high accuracy and efficiency, where the optimization criterion was least-squares (LS) and the deformation model was affine. This previous algorithm was limited to intra-modal registration because of the properties of the LS criterion. In this paper, rather than LS, we choose a new criterion named Mutual Information (MI) which appears to have a much greater potential for inter-modal registration and image fusion. This new criterion, introduced independently in [3] and [10], has been proposed recently in the context of medical image registration [2] as well as of computer vision [10].

Overall, MI is a measure of statistical *dependency* between two data sets [6]. In applications intended to fuse two images, this measure is intuitively more appropriate than the previous measure of statistical *correlation* offered by LS. To understand why, consider a pair of images from different modalities that are already in perfect alignment: at many corresponding geometric locations in the two images, the associated gray-levels will exhibit a consistent mapping. In other words, when correctly co-registered, it is likely that there will exist some kind of a global relation, or mapping, between the gray-levels of one of the images, and those of the other, whatever this mapping may be (one-to-one, or even one-to-few or few-to-one). At the same time, the correlation between the gray-levels may happen to be low or even negative. Even in this case however, the statistical dependency is strong and the MI measure still yields a high value.

Suppose now that we move out of registration by geometrically deforming one of the images: the mapping of the gray-levels will degrade, *i.e.* will become less consistent, and MI will diminish. In the extreme case where the two images become completely independent of one another, MI will take a zero value, a condition signifying that there is no way to predict any part of one image from the information stored in the other. At the extreme opposite, seeking the geometric deformation that maximizes MI is a way to perform registration.

In this paper, we investigate the interplay of MI with our previous registration scheme, which used an explicit continuous modeling of the image that was based on splines, and which took advantage of the concept of multi-resolution. For the estimation of MI itself, we also explore techniques that tend to reduce the number of its local maxima, with the goal of simplifying the optimization process.

2. DEFINITIONS

Let $f_{\mathrm{T}}(\mathbf{x})$ and $f_{\mathrm{R}}(\mathbf{x})$, $\forall \mathbf{x} \in \mathbf{R}^n$, be a test and a reference image, respectively. Let $T\{f(\mathbf{x})\} = f(\mathbf{A}\mathbf{x} - \mathbf{b})$ be an affine transformation that we want to apply to the test image $f_{\mathrm{T}}(\mathbf{x})$. Let $V \subseteq \mathbf{R}^n$ be a region of interest associated with $f_{\mathrm{R}}(\mathbf{x})$. Let $\mathbf{L}_{\mathrm{T}} \subseteq \mathbf{R}$ and $\mathbf{L}_{\mathrm{R}} \subseteq \mathbf{R}$ be a test and a reference discrete set of intensities, respectively. Finally, let $w(\xi) \geq 0$, $\forall \xi \in \mathbf{R}$, be a Parzen window.

We proceed as follows: we compute and sample the joint Parzen histogram, $\forall \iota \in \mathbf{L}_{\mathrm{T}}$, $\forall \kappa \in \mathbf{L}_{\mathrm{R}}$,

$$h(\iota, \kappa) = \alpha \sum_{\mathbf{x} \in V} w(\iota - f_{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{b}))\, w(\kappa - f_{\mathrm{R}}(\mathbf{x})), \tag{1}$$

where α is selected such that

$$\sum_{\iota \in \mathbf{L}_{\mathrm{T}}} \sum_{\kappa \in \mathbf{L}_{\mathrm{R}}} h(\iota, \kappa) = 1. \tag{2}$$

We compute the marginal histograms

$$h_{\mathrm{T}}(\iota) = \sum_{\kappa \in \mathbf{L}_{\mathrm{R}}} h(\iota, \kappa), \ \forall \iota \in \mathbf{L}_{\mathrm{T}}, \tag{3}$$

and

$$h_{\mathrm{R}}(\kappa) = \sum_{\iota \in \mathbf{L}_{\mathrm{T}}} h(\iota, \kappa), \ \forall \kappa \in \mathbf{L}_{\mathrm{R}}. \tag{4}$$

The goal of the registration problem is to find the set of transformation parameters $\{\hat{\mathbf{A}}, \hat{\mathbf{b}}\}$ that optimizes (maximizes) the following criterion:

$$S = \sum_{\iota \in \mathbf{L}_{\mathrm{T}}} \sum_{\kappa \in \mathbf{L}_{\mathrm{R}}} h(\iota, \kappa) \log\left(\frac{h(\iota, \kappa)}{h_{\mathrm{T}}(\iota) h_{\mathrm{R}}(\kappa)}\right). \tag{5}$$

This last expression (the mutual information) is valid only when the sets $\mathbf{L}_{\mathrm{T}}$ and $\mathbf{L}_{\mathrm{R}}$ exclude those values ι and κ for which $h_{\mathrm{T}}(\iota) = 0$, $h_{\mathrm{R}}(\kappa) = 0$ and $h(\iota, \kappa) = 0$, respectively. Note that they do not contribute to the criterion S, because $\lim(0^{+} \log(0^{+})) = 0$.

3. OPTIMIZATION

As is well-known [4, 7], the use of a multi-resolution scheme has at least one important benefit: dealing with a low-resolution (coarse) image translates into dealing with a low-resolution (smoothed) criterion. In turn, this helps to avoid getting trapped into some local optima (because they do not exist anymore at low resolution), and often widens the capture area of the true optimum. A side effect is an important improvement in speed, provided three conditions are met: 1) use of a robust algorithm at the coarsest level, 2) at intermediate levels, avoidance of an over-accurate registration, and 3) at the finest level, use of an optimization algorithm that is super-linear in convergence.

In this paper, we use the unconstrained optimization method of Jeeves [1] in order to find iteratively the maximum of S. It does not require gradient estimations but solely function evaluations. Its working principle is to explore the close neighborhood of the actual guess, one step at a time, in a star-fashion along the coordinate axis, and to climb S at each opportunity. The step-size is tuned by an adaptation mechanism based on the history of the successes and failures encountered during the optimization procedure.

This method is an all-purpose optimization scheme that does not capitalize on properties such as a quadratic form for the optimum. It is generally not considered to be the most efficient one available; however, its simplicity makes it a good candidate for the exploration of a new criterion. A better search technique would possibly make use of gradient information; since we introduced continuity by the way of Parzen windowing, even though $\mathbb{L}_{\mathrm{T}}$ and $\mathbb{L}_{\mathrm{R}}$ are discrete, the quantity $\partial S / \partial\{\mathbf{A},\mathbf{b}\} = \partial S / \partial\lambda$ is easy to compute:

$$\frac{\partial S}{\partial\lambda} = \sum_{\iota\in\mathbb{L}_{\mathrm{T}}} \sum_{\kappa\in\mathbb{L}_{\mathrm{R}}} \left[\frac{\partial h(\iota,\kappa)}{\partial\lambda} \left(1 + \log(\frac{h(\iota,\kappa)}{h_{\mathrm{T}}(\iota) h_{\mathrm{R}}(\kappa)}) \right) - h(\iota,\kappa) \left(\frac{\partial h_{\mathrm{T}}(\iota)/\partial\lambda}{h_{\mathrm{T}}(\iota)} + \frac{\partial h_{\mathrm{R}}(\kappa)/\partial\lambda}{h_{\mathrm{R}}(\kappa)} \right) \right],$$

where

$$\frac{\partial h_{\mathrm{T}}(\iota)}{\partial\lambda} = \sum_{\nu\in\mathbb{L}_{\mathrm{R}}} \frac{\partial h(\iota,\nu)}{\partial\lambda}, \quad \frac{\partial h_{\mathrm{R}}(\kappa)}{\partial\lambda} = \sum_{\nu\in\mathbb{L}_{\mathrm{T}}} \frac{\partial h(\nu,\kappa)}{\partial\lambda},$$

where

$$\frac{\partial h(\iota,\kappa)}{\partial\lambda} = \alpha \sum_{\mathbf{x}\in V} \zeta(\mathbf{x})\, w(\kappa - f_{\mathrm{R}}(\mathbf{x})) \left. \frac{\partial w(\xi)}{\partial\xi} \right|_{\xi=\iota-f_{\mathrm{T}}(\mathbf{A}\mathbf{x}-\mathbf{b})}$$

and

$$\zeta(\mathbf{x}) = \left(\left. \frac{-\mathrm{d}f_{\mathrm{T}}(\mathbf{t})}{\mathrm{d}\mathbf{t}} \right|_{\mathbf{t}=\mathbf{A}\mathbf{x}-\mathbf{b}} \right)^{\mathrm{T}} \frac{\partial}{\partial\lambda} \left(\mathbf{A}\mathbf{x} - \mathbf{b} \right).$$

4. INTERPOLATION

The interpolation process is a crucial part of any registration scheme because it uniquely defines the behavior of the transformation when the samples do not fall on a grid, a condition generally true in a sub-pixel case. In addition to this general consideration, interpolation requires special attention in the case of MI because this process introduces new gray-levels that may interfere with the computation of (1). On one hand, it is desirable to have a high geometric precision; for example, sinc (Fourier) interpolation satisfies exact reversibility for a translation. On the other hand, one would prefer not to introduce improper gray-levels; for example, the ringing associated with sinc interpolation may significantly corrupt the estimation of MI.

As an illustration of the potential sensitivity of MI to spurious values, consider a 10×10 patch of data centered on a step function (2 levels). The linear interpolation of this image after a very small sub-pixel translation across the edge introduces 10 new members for a third gray-level. Ideally, we would like that the entropy (a pivotal component of MI, see [6]) remains unaffected by this small translation. However, before translation this term was $H\{f(\mathbf{x})\} = 1$. After translation, we have $H\{T\{f(\mathbf{x})\}\} = 1.03 + 0.33$, the last value being the contribution of the new gray-level (whatever it may be), which represents only 10% of all pixels.

We present in this paper a series of experiments that tend to show that it is beneficial to pay more attention to geometry by using higher interpolation models, and that the corruption of the mutual information that arises when the interpolation process artificially creates new gray-levels is in fact immaterial. Our general procedure for interpolation will be to fit a separable B-spline model to the data, apply to the model whatever transformation is required, and then resample the transformed model [8, 9].

5. MULTI-RESOLUTION

In [7] we introduced an efficient optimization scheme that used a coarse-to-fine iterative refinement strategy. Substituting MI for LS, our task is now to show that the new criterion is well behaved in the context of multi-resolution. An extra twist comes from the estimation of the joint histogram: at any resolution level l, an image of size $\mathrm{N}_l \times \mathrm{N}_l$ contains at most N_l^2 different values. In absence of any *a priori* knowledge on the test and reference images, and given that the joint histogram $h(\iota,\kappa)$ holds $\mathrm{I}^2 = \mathrm{Card}(\mathbb{L}_{\mathrm{T}} \times \mathbb{L}_{\mathrm{R}})$ independent values, it certainly does not make sense to have $\mathrm{I}^2 > \mathrm{N}_l^2$. This sparsity condition imposes a reasonable upper bound to the number of gray-levels I we may introduce at each resolution level l. On the other hand, we have the absolute, l-independent, lower bound $\mathrm{I} \geq 2$. We present in this paper some experimental evidence that an optimal I_l may indeed be found between these two bounds.

6. BINNING

A relevant consideration in our framework is the choice of the sets $\mathbb{L}_{\mathrm{T}}$ or $\mathbb{L}_{\mathrm{R}}$. While any quantization scheme is admissible, be it regular or irregular, we prefer to consider linear quantization only, which alleviates complications that would otherwise arise in the application of Parzen windowing. At each level l, we determine the extrema of $f(\mathbf{x})$ and simply divide this range into I quantization intervals. The Parzen window we select is the centered, scaled, B-spline of order n given by $w(x) = \beta^{(n)}(a x)$. When $a = 1$, this particular Parzen window has integral unity $\forall n$, and the sum of its discrete samples is also unity, which are desirable properties for Parzen windows.

This strategy does not preclude preprocessing of the image; for example, it is easy to show that, in a continuous case, MI is invariant under strictly monotonic transformations of gray-levels (*e.g.*, histogram equalization). In our context, we interpret such a transformation as a way to have $\mathbb{L}$ perform irregular sampling.

7. ILLUSTRATION OF MI

In Figure 1, we show the "mandrill" color image at left, and at right the same image with respect to the cyan, yellow and black channels, but with the magenta channel rotated around its center by $10°$ and displaced by $\Delta x = 3$ and $\Delta y = 5$. In Figure 2, the corresponding joint histograms of the magenta and black channels show that the transformation smears h. Figure 3 shows the relevant individual channels.

While the complete criterion S given in (5) includes some terms other than (or derived from) $h(\iota,\kappa)$, image fusion based on mutual information often tends to select the transformation that produces the sharpest $h(\iota,\kappa)$. One can find in [5] a method that tries to directly maximize the degree of clustering in the intensity space, without resorting to MI.

8. EXPERIMENTS

We take as multi-modal data the magenta and black channel of the CMYK representation of a color image (Figure 3). The benefit of this choice is that we have knowledge of the ground-truth transformation: identity has them in perfect registration. When a non-trivial transformation is applied, however, we are careful to mask out the irrelevant parts (those that would be in need of extrapolation).

8.1. Shape and size of the Parzen window

As preliminary experiment, we look at the influence of the shape of the Parzen window. Using a B-spline $w(x)=\beta^{(n)}(ax)$, we consider two parameters: its order n, and its knot spacing $\mathrm{w}=1/a$. Figure 4 gives S as a function of a translation (in pixels) of the M-channel with respect to the K-channel. As w increases, each Parzen window covers more bins, which produces a smoothing of S. This widens the capture area for the optimum, perhaps at the cost of a loss in accuracy due in part to a local flattening of the curve near the optimum, and in part to a global decrease of the dynamic range. To a lesser extent, this is also true when the order of the spline increases: for a fixed w, a higher-order spline generates more overlap than a lower-order spline.

After examination of Figure 4, a third order spline ($\beta^{(3)}$) with natural sampling ($\mathrm{w}=1$) seems to be an appropriate trade-off between the avoidance of the grid effect, particularly pronounced when the overlap is insufficient, and the loss of dynamic range experienced when the overlap is too important. In this experiment, we used $\mathrm{I}=16$ bins, third-order interpolation and images of size 256×256.

8.2. Number of bins

The set of experiments we are about to present is organized around the following procedure: 1) pick an affine transformation T at random and apply it to f_{R} while keeping f_{T} fixed; 2) using an iterative scheme, start with the identity as a first guess for the initial transformation and register f_{T} to the already transformed f_{R}, which yields an estimated transformation $\hat{T}$; 3) compare T and $\hat{T}$ by computing some quality measure Q. Our choice for this quality measure is the average registration error on a pixel per pixel basis; it is given by the sum over $\mathbf{x}\in V$ of the distance between $T(\mathbf{x})$ and $\hat{T}(\mathbf{x})$,

$$Q=\frac{1}{\mathrm{Card}(V)}\sum_{\mathbf{x}\in V}\left\|T(\mathbf{x})-\hat{T}(\mathbf{x})\right\|. \qquad (6)$$

Trying to empirically determine I_l, the best number of bins at a given resolution level l, we conduct a series of 50 trials per experiment; for each trial we generate a transformation consisting of a random rotation with $|\theta|\leq\pi/36$, and a random translation with $(|\Delta x|,|\Delta y|)\leq(5,5)$. Note that the estimated transformation $\hat{T}$ is affine in general, and may depart from a pure rotation. We compute a resolution pyramid containing l levels, and perform optimization on the coarsest level only (the finest level $l=1$ has a size 256×256). We then propagate the resulting transformation up to the finest level where the quality measure Q is estimated. For this experiment, we keep using natural sampling for the Parzen window, and we consider $l=3$, which corresponds to a fourfold magnification in the average registration error. Each data point gives the raw (not scaled down) average registration error, pooled over those of the 50 random trials that converged close enough to the correct solution. There were very few outliers, and those were obvious to detect; out of the 2800 trials conducted for establishing Figures 5 and 6, we had only 216 outliers, which were primarily a result of using too few bins.

Figure 5 shows the result of these experiments when bilinear interpolation is used for the computation of the affine transformation. The general trend is as expected: a less than optimal result is obtained when there are too many bins, because $h(\iota,\kappa)$ is poorly estimated, and a bad result when there are not enough bins, because $h(\iota,\kappa)$ is not discriminant enough. In between, an optimal number of bins I_l can be selected; according to Figure 5, we have $\mathrm{I}_3\approx10$. We note also that a Parzen window $w=\beta^{(3)}$ often does better than $w=\beta^{(1)}$, which is consistent with the conclusions of the previous experiment 8.1.

Figure 6 shows the same experiment as in Figure 5, where bilinear interpolation has been substituted by bicubic interpolation. The use of this higher quality model results in an improvement in accuracy (compare to Figure 5), especially when the number of bins is small. Another benefit is a decrease in the value of the optimal number of bins; in this case, we have $\mathrm{I}_3\approx6$. This may become a significant factor for optimization algorithms making use of the gradient of S, since this latter involves a triple summation over $\mathbf{L}$.

9. CONCLUSIONS

In this paper, we have investigated the role of mutual information as a new criterion for image registration. We have formulated an expression of the criterion that is based on a discrete binning process, but that is continuous with respect to the transformation parameters (hence derivable) since it involves Parzen windows. We have shown experimentally that the shape of the Parzen window, especially its width, is a critical parameter. In the context of a multi-resolution approach, we have developed arguments hinting at the dependence of the number of bins on the resolution level. For a given level, we have shown experimental evidence of the existence of such an optimum. In addition, we have determined that it is beneficial to use high-order interpolation schemes, even though they interfere with the estimation of the mutual information by introducing spurious gray-levels.

The overall accuracy obtained by this new criterion is satisfying, especially in the context of multi-modal data. Even when we propagate to the finest resolution a transformation estimated on the third level of a resolution pyramid, hence multiplying registration errors by four (and at the same time using only a sixteenth of the available data), we get an accuracy at the finest level that is still largely sub-pixel. These encouraging results drive us now to implement the whole multi-resolution approach, incorporating more efficient optimization schemes along the way.

BIBLIOGRAPHY

[1] I.N. Bronshtein and K.A. Semendyayev, *Handbook of Mathematics*. Verlag Harri Deutsch, 1985.

[2] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens and G. Marchal, "Automated Multi-Modality Image Registration Based on Information Theory," in *Information Processing in Medical Imaging*, Y. Bizais, Ed., pp. 263–274, 1995.

[3] A. Collignon, D. Vandermeulen, P. Suetens and G. Marchal, "3D Multi-Modality Medical Image Registration Using Feature Space Clustering," in Proc. *Computer Vision, Virtual Reality, and Robotics in Medicine*, Nice, France, April, 1995, pp. 195–204.

[4] J.-P. Djamdji, B. Albert and R. Manière, "Geometrical Registration of Images: The Multiresolution Approach," *Photogrammetric Engineering & Remote Sensing*, vol. 59, no. 5, pp. 645–653, May 1993.

[5] D.L.G. Hill, C. Studholme and D.J. Hawkes, "Voxel Similarity Measures for Automated Image Registration," in Proc. *SPIE Proceedings of the Third Conference on Visualization in Biomedical Computing*, 1994, pp. 205–216.

[6] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Inc., 1991.

[7] P. Thévenaz, U.E. Ruttimann and M. Unser, "Iterative Multi-Scale Registration without Landmarks," in Proc. *International Conference on Image Processing*,

Washington, D.C., U.S.A., October 23–26, 1995, vol. III, of 3, pp. 228–231.

[8] M. Unser, A. Aldroubi and M. Eden, "B-Spline Signal Processing: Part I—Theory," *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 821–832, February 1993.

[9] M. Unser, A. Aldroubi and M. Eden, "B-Spline Signal Processing: Part II—Efficient Design and Applications," *IEEE Transactions on Signal Processing*, vol. 41, no. 2, pp. 834–848, February 1993.

[10] P. Viola and I. Wells, William M., "Alignment by Maximization of Mutual Information," in Proc. *International Conference on Computer Vision*, Boston, MA, USA, June, 1995, pp. 16–23.

FIGURES

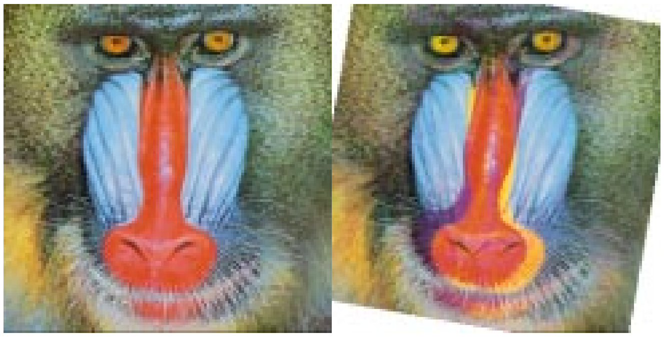

Figure 1: Color image (left) and its corrupted version (right), where the magenta channel has been rotated and displaced.

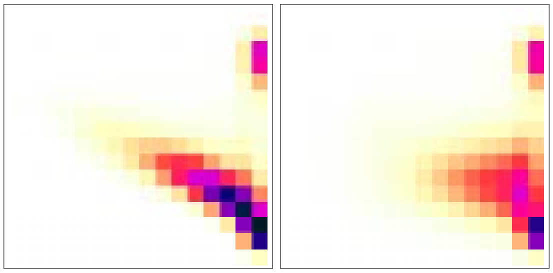

Figure 2: Joint histograms of the black and magenta channels of Figure 1. Left: registered, right: out of registration.

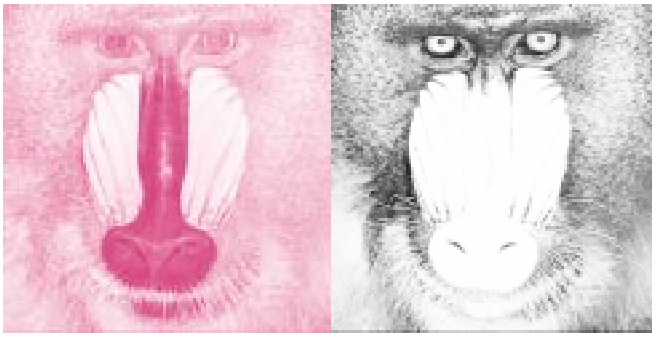

Figure 3: Magenta (left) and black (right) channel of the "mandrill" color image, used as test and reference image, respectively.

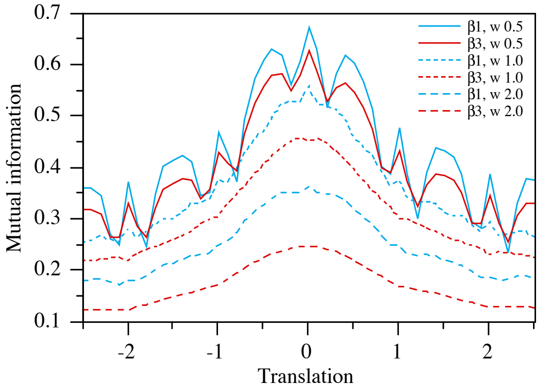

Figure 4: Dependence of the mutual information on the Parzen window type and size.

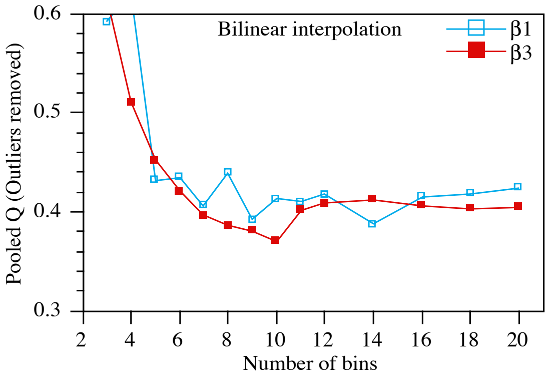

Figure 5: Dependence of the quality measure on the number of bins with bilinear interpolation.

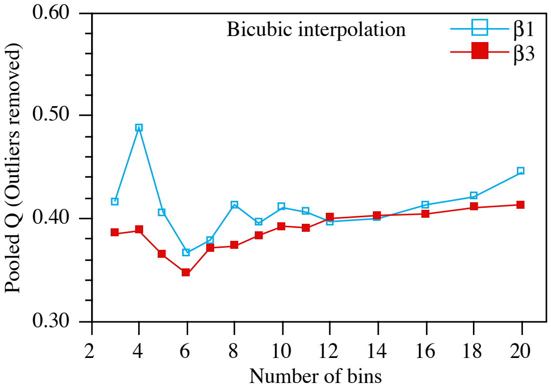

Figure 6: Dependence of the quality measure on the number of bins with bicubic interpolation.