# 6

# Post-Processing of PIV Data

The recording and evaluation of PIV images has been described in the previous two chapters. Investigations employing the PIV technique usually result in a great number of images which must be further processed. If looking for statistical quantities the recorded data can easily amount to some gigabytes, which is now possible with today's computer hardware. Even more data per investigation are to be expected in future. Thus, it is quite obvious that a fast, reliable and fully automatic post-processing of the PIV data is essential.

In principle, post-processing of PIV data is characterized by the following steps:

**Validation of the raw data.** After automatic evaluation of the PIV recordings, a certain number of obviously incorrectly determined velocity vectors (outliers) can usually be found by visual inspection of the raw data. In order to detect these incorrect data, the raw flow field data have to be validated. For this purpose, special algorithms have to be developed, which must work automatically.

**Replacement of incorrect data.** For most post-processing algorithms (e.g. calculation of vector operators) it is required to have complete data fields as is quite naturally the case for numerically obtained data. Such algorithms will not work if gaps (data drop-outs) are present in the experimental data. Thus, means to fill the gaps in the experimental data must be developed.

**Data reduction.** It is quite difficult to inspect several hundred velocity vector maps and to describe their fluid mechanical features. Usually techniques like averaging (in order to extract the information about the mean flow and its fluctuations), conditional sampling (in order to distinguish between periodic and non-periodic parts of the flow) and vector field operators (e.g. vorticity, divergence in order to detect structures in the flow) are applied.

**Analysis of the information.** At present this is the most challenging task for the user of the PIV, technique. PIV being the first technique to offer information about complete instantaneous velocity vector fields, allows new insights in old and new problems of fluid mechanics. Tools for analysis such as proper orthogonal decomposition (POD) [276] or neural networks [250] are applied to PIV data.

**Presentation and animation of the information.** A number of software packages – commercial as well as in-house developed ones – are available for the graphical presentation of the PIV field data. It is also very important to support the easier understanding of a human observer of the main features of the flow field. This can be done by contour plotting, color coding, etc. Animation of the PIV data is very useful for better understanding in the case of time series of PIV recordings or 3D data.

In the following sections those steps of post-processing with special requirements due to the PIV technique will be explained in greater detail.

## 6.1 Data Validation

Some of the problems associated with PIV raw data after automatic evaluation can be seen in figure 6.1. It shows the instantaneous flow field above a NACA 0012 airfoil at a free stream Mach number Ma = 0.75. The vector of the average flow velocity (344 m/s) as calculated for this PIV recording has been subtracted from each individual velocity vector in order to enhance details of the flow field. The supersonic flow regime above the leading edge of the airfoil and the terminating shock with its strong velocity gradient can clearly be detected.
For more details on the experiments see chapter 9. Typical features of incorrect velocity vectors, which can be detected in figure 6.1, are that:

- their magnitude and direction differ considerably from their surrounding neighbors,

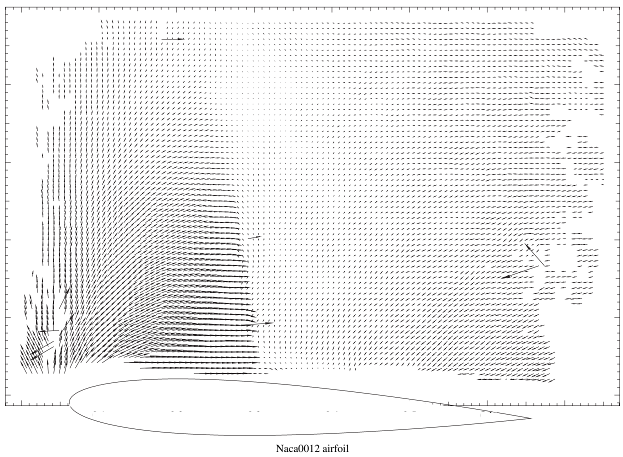


**Fig. 6.1.** Velocity vector map (raw data) of instantaneous flow field $(U-\overline{U},V)$ above NACA 0012 airfoil (Ma= 0.75, $\alpha=5^\circ$, $l_\mathrm{c}=20\,\mathrm{cm}$, $\tau=4\,\mu\mathrm{s}$, $\overline{U}=344\,\mathrm{m/s}$).

- they very often appear at the edges of the data field (near the surface of the model, at the edges of drop-out areas, at the edges of the illuminated area),
- in most cases, they appear as single incorrect vectors.

From this description we can conclude that it is most likely that during the evaluation procedure a correlation peak was detected which is due to noise or artifacts (model surface, noise of different sources, etc.) and not due to the correlation of properly matched image pairs. These questionable or spurious data points are frequently defined as *outliers*. In general an outlier may be defined as an observation (data point) which is very different from the rest of the data based on some measure.

The human perceiver is very efficient in detecting these outliers. For a small number of PIV recordings these erroneous velocity vectors may be treated interactively. This is no longer possible if a great number of recordings has to be evaluated. However, for the further processing of the flow field data it is absolutely necessary to eliminate all such erroneous data. All subsequent operations involving differential operators on the raw vector data still including such outliers would enhance and smear out these errors locally and could thus mask data of good quality. Differential operators are, for example, the divergence, the vorticity operator or the calculation of differences between numerical and experimental flow field data. In contrast to this, the application of operators utilizing averaging processes over a great number of data, mean value, variance, degree of turbulence, etc., are less severely affected by a few incorrect data values. It follows that all PIV data should generally be checked for erroneous data. Because of the great amount of data this can only be performed by means of an automatic algorithm. The guiding principle for handling questionable data should be:

- The algorithm must ensure with a high level of confidence so that no questionable data are stored in the final PIV data set.
- Questionable data should be rejected, if it cannot be decided by application of the algorithm whether data are valid or not.

As a consequence of the application of the validation algorithm, the number of PIV data obtained from a given recording will be reduced by approximately 0.1–1.5% depending on the quality of the PIV recording and the type of flow to be investigated. The problem of filling gaps where no valid data were found on the flow field map (by means of interpolation or extrapolation) should only be performed after completing the data validation. It should again be emphasized that this procedure prevents information arising from incorrect data being spread into areas with data of good quality. For the same reason no smoothing of data should be carried out before validating the data.

The challenge in data validation is to provide algorithms that strike a good balance between overdetection, that is the removal of valid data, and underdetection in which too many spurious vectors are accepted. Different techniques for PIV data validation have been described in the literature [115, 119, 258, 273, 274]. However, to date no general solution can be offered for the problem of data validation in PIV. However, some degree of generalized validation is possible through the use of the normalized median filter [275] (see page 185). Variable threshold approaches determine the detection threshold from filtered versions of the unvalidated data set [264, 268].

In our applications we use several algorithms, which have been developed and tested utilizing real PIV recordings obtained in different experimental situations and for different types of flows [266]. Some of these modules are alternatives, some of them may be applied successively. Two of these modules, the *global histogram operator* and the *dynamic mean value operator*, will be discussed in detail in the following. Further validation schemes are described as well.

Some definitions will be made for the following discussion of the different data validation algorithms. The instantaneous velocity vector field $(U, V)$ has been sampled ("interrogated") at positions which form a regular grid in the flow field. In our case the grid, a part of which is shown in figure 6.2, consists of $I \times J$ grid points in the $X$ and $Y$ directions with constant distance $\Delta X_{\rm step}, \Delta Y_{\rm step}$ between neighboring grid points in both directions. The two-dimensional velocity vector at the position $i, j$ $(i = 1 \cdots I, j = 1 \cdots J)$ is denoted $\mathbf{U}_{\rm 2D}(i, j)$. In the following discussion the relation between the central velocity vector $\mathbf{U}_{\rm 2D}(i, j)$ and one of its nearest neighbors $\mathbf{U}_{\rm 2D}(n)$ is considered. The nearest neighbors are labeled by $n, (n = 1, ..., N)$. Usually, $N$ is chosen to be eight. The distance $d$ between the central velocity vector $\mathbf{U}_{\rm 2D}(i, j)$ and its nearest neighbors $\mathbf{U}_{\rm 2D}(n)$ is either $\Delta X_{\rm step}, \Delta Y_{\rm step}$ or $\sqrt{\Delta X^2_{\rm step} + \Delta Y^2_{\rm step}}$, depending on its position on the grid. The magnitude of the vector difference between the central velocity vector $\mathbf{U}_{\rm 2D}(i, j)$ and $\mathbf{U}_{\rm 2D}(n)$ is $|\mathbf{U}_{{\rm diff},n}| = |\mathbf{U}_{\rm 2D}(n) - \mathbf{U}_{\rm 2D}(i, j)|$.
For the demonstration of the performance of the different algorithms for data validation, the flow field shown in figure 6.3 is utilized. It represents the lower left part of the flow field already shown in figure 6.1.

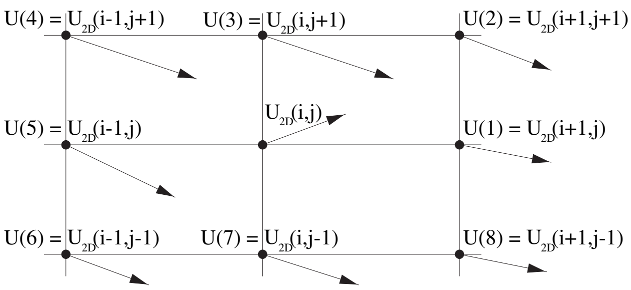


**Fig. 6.2.** Sketch of data grid with notation of vectors.

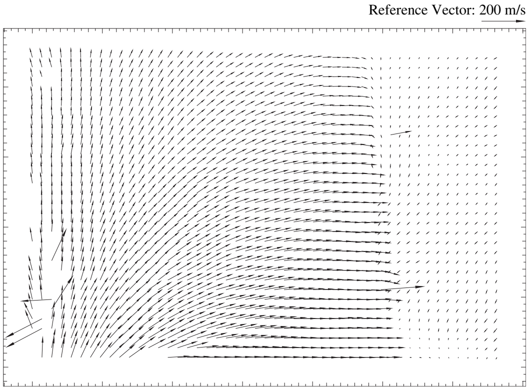


**Fig. 6.3.** Enlarged area (lower left part) of figure 6.1 for the demonstration of the data validation algorithm.

### 6.1.1 Global Histogram Operator

**Principle**. It is assumed that for real flow fields the vector difference between the neighboring velocity vectors is smaller than a certain threshold. This is certainly true as long as the length scale of the flow is much greater than the distance $d$ between the position of neighboring vectors. Consequently all correct velocity vectors must lie within a continuous area in the $(u, v)$ plane for flow fields without discontinuities. As a first step this criterion is used to discriminate against incorrect data.

**Procedure**. A 2D histogram of the displacement (or velocity) data is obtained by plotting the positions of all recovered correlation peaks in a common correlation plane. An equivalent presentation is given in figure 6.4 where a point has been plotted in the correlation plane at the location of the highest correlation peak for each interrogation window.

Two separated areas of accumulated correlation peaks (velocity vectors) can be detected, one circular region (I) and a second region (II) with greater scattering of the peak's locations (i.e. of the velocity vectors). Now a rectangular box can automatically be calculated, circumscribing the area with the highest accumulation of displacement peaks (velocity vectors). In the next step of the validation procedure the displacement peaks detected for each interrogation window will be checked individually. All displacement peaks (velocity vectors) lying outside the rectangle(s) or other suitable boundary will be marked and rejected.

**Discussion**. Most outliers due to noise in the correlation plane can be rejected by application of this simple algorithm. Also, if autocorrelation is employed, data originating from the vicinity of the origin of the correlation plane (zero order central peak) can be eliminated by defining another rectangle of rejection around the zero order peak (not shown in figure 6.4). Noise originating from the recording process, which is concentrated in certain spectral bands (noise from AC sources, imperfections of the optics for interrogation,

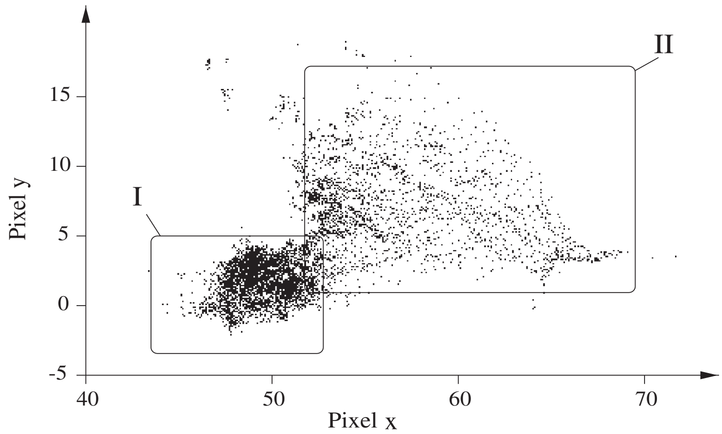


**Fig. 6.4.** Location of correlation peaks in the correlation plane. Rectangles indicate areas of plausible data, area (I) Ma < 1, (II) Ma > 1.

structures of the CCD-camera, etc.) can be detected and eliminated because in most cases its response is located in areas of the correlation plane which are not connected to those areas where the correct velocity vectors accumulate.

Figure 6.4 also demonstrates that there may be situations with two or more areas in the correlation (or velocity) plane, where the peaks (or velocity vectors) accumulate. This is the case if discontinuities are present in the flow field, for example for transonic flow fields, if shocks are embedded in the flow field. In figure 6.4 the area marked (I) is due to the subsonic part of the flow field at the right side of figure 6.3, whereas the area (II) is due to the supersonic part of the flow field just above the leading edge of the airfoil. Summarizing, one can say that the global histogram operator employs physical arguments (upper and lower limit of possible flow velocities) to remove all data, which physically cannot exist in the flow field. Moreover, the inspection of the global velocity histogram gives useful information about the quality of the PIV evaluation (number of incorrect data due to noise, dynamic range of the flow field, maximum utilization of the optimal range for PIV evaluation by selecting the proper time delay between the light pulses for illumination, etc.). If at this stage of the validation process a velocity vector is rejected, it will automatically be checked whether for this grid position the automatic evaluation procedure has detected more than one peak in the correlation plane (see section 5.4.5). If this should be the case, these data are checked as well, whether they fulfill the criteria explained above. It is obvious that the maximum number of peaks admitted during the automatic evaluation for a fixed grid position should be limited to two or three, as otherwise – with a great number of peaks for selection – the chance would be rather high to pick up peaks due to noise, which, however, accidentally fulfill the criterion for selection.

### 6.1.2 Dynamic Mean Value Operator

**Principle**. Many PIV validation schemes described in the literature makes use of mean value tests. These algorithms check each velocity vector individually by comparing its magnitude $|\mathbf{U}_{2D}(i,j)|$ with the average value over its nearest neighbors $\boldsymbol{\mu}_{\mathbf{U}}(i,j)$. Typically, a $3 \times 3$ neighborhood with eight nearest neighbors is selected. The velocity vector to be validated will be rejected if the absolute difference between its magnitude and the average over its neighbors is above a certain threshold $\epsilon_{\text{thresh}}$. The test can be modified by applying it not only to the magnitude but also to the $U$ and $V$ components of the vector, or by utilizing a larger number of neighbors for comparison.

However, the application of this test to transonic flows has shown some problems if shocks are present in the flow field (discontinuity of the flow velocity along a line). Thus, the algorithm had to be improved for flows with velocity gradients by locally varying the threshold level $\epsilon$ for validation.

**Procedure**. The following expression is calculated for $N = 8$ closest neighbors:

$$\boldsymbol{\mu}_{\mathbf{U}}(i,j) = \frac{1}{N}\sum_{n=1}^{N} \mathbf{U}_{2D}(n) \; . \tag{6.1}$$

The averaged magnitude of the vector difference between the average vector and its 8 neighbors is also calculated:

$$\sigma^2_{\mathbf{U}}(i,j) = \frac{1}{N}\sum_{n=1}^{N} \left(\boldsymbol{\mu}_{\mathbf{U}}(i,j) - \mathbf{U}_{2D}(n)\right)^2 \; . \tag{6.2}$$

The criterion for data validation is:

$$|\boldsymbol{\mu}_{\mathbf{U}}(i,j) - \mathbf{U}_{2D}(i,j)| < \epsilon_{\text{thresh}} \tag{6.3}$$

where $\epsilon_{\text{thresh}} = C_1 + C_2\,\sigma_{\mathbf{U}}(i,j)$ with $C_1, C_2 =$ constants.

Problems will arise in the case of drop-outs within the data field or at the edges of the data field, that is if less than $N = 8$ neighbors are available for comparison. Satisfying results were obtained when the drop-out areas were filled with the mean value of the whole data field and lines and rows were added at each edge of the data field by doubling the outermost lines and rows. However, these artificially generated data were only kept during the application of the local mean value test or other validation tests.

**Discussion**. The effect of the application of the dynamic mean value operator is demonstrated in figure 6.5, where all velocity vectors identified as "incorrect" are marked by thick vector symbols. Figure 6.5 clearly shows the power of the algorithm described above as those velocity data, which are incorrect – as they are obviously different in magnitude and direction from their neighbors – have been consequently marked. However, the application of this algorithm does not lead to difficulties with the strong velocity gradients across

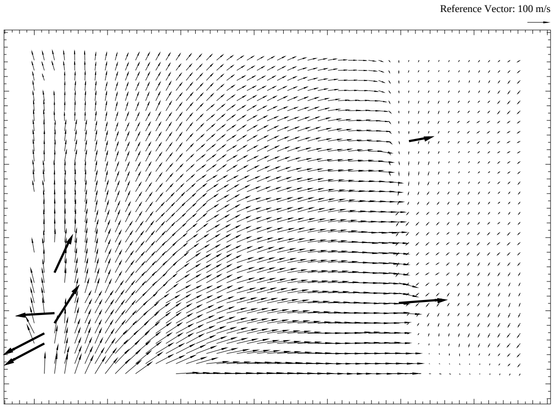


**Fig. 6.5.** Application of the dynamic mean value operator on the velocity vector map of figure 6.3. Vectors identified as "incorrect" are marked by thick vector symbols.

the shock, that is no correct data are marked as "incorrect". Thus, it has been shown that this algorithm is able to handle flows with strong velocity gradients by the introduction of the locally varying threshold level for validation (see figure 6.6). The constants $C_1$, $C_2$ have to be determined once experimentally and can then be utilized for the whole series of PIV recordings taken for the same type of flow.

### 6.1.3 Vector Difference Test

Similar to the previous filter the gradient filter or vector difference filter computes the magnitude of the vector difference of a particular vector in question

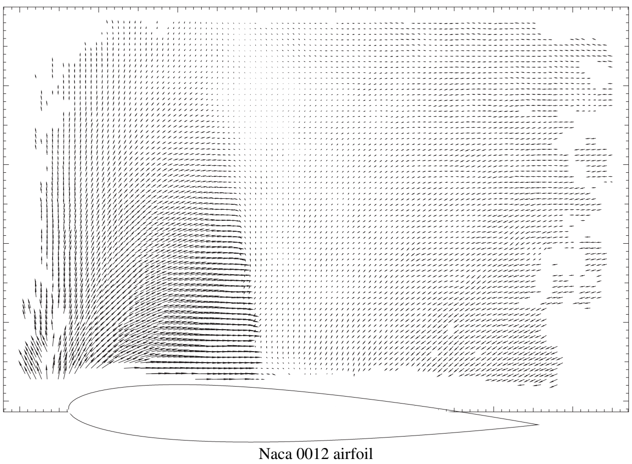


**Fig. 6.6.** Velocity vector map of the instantaneous flow field above a NACA 0012 airfoil of figure 6.1 after clean-up.

$\mathbf{U}_{2D}(i,j)$ to each of its four or eight neighbors $\mathbf{U}_{2D}(n)$:

$$|\mathbf{U}_{\text{diff},n}| = |\mathbf{U}_{2D}(n) - \mathbf{U}_{2D}(i,j)| < \epsilon_{\text{thresh}} \quad \text{with} \quad \epsilon_{\text{thresh}} > 0\ . \tag{6.4}$$

Rather than averaging the vector differences as for the dynamic mean filter the idea is to count the number of instances for which the validation criterion $|\mathbf{U}_{\text{diff},n}| < \epsilon_{\text{thresh}}$ is violated. A displacement vector can then be classified as questionable when it is 'conflicting' with at least half its neighbors.

### 6.1.4 Median Test

PIV data validation by means of median filtering has been proposed by Westerweel [274]. While median filtering is frequently utilized in image processing to remove spurious noise, it may also by used for the efficient treatment of spurious velocity vectors. Median filtering simply speaking means that all neighboring velocity vectors $\mathbf{U}_{2D}(n)$ are sorted linearly either with respect to the magnitude of the velocity vector, or their $U$ and $V$ components. The central value in this order (i.e. either the fourth or fifth of eight neighbors) is the median value. The velocity vector under inspection $\mathbf{U}_{2D}(i,j)$ is considered valid if

$$|\mathbf{U}_{2D}(\text{med}) - \mathbf{U}_{2D}(i,j)| < \epsilon_{\text{thresh}}\ .$$

### 6.1.5 Normalized Median Test

A slight modification of the median test results in a very powerful validation scheme for spurious velocity vectors. Westerweel & Scarano [275] demonstrated that a normalization of the standard median test given in equation (6.1.4) yields a rather universal probability density function for the residual such that a single threshold value can be applied to effectively detect spurious vectors. The normalization requires that the residual $r_i$, defined as: $r_i = |\mathbf{U}_i - \mathbf{U}_{\text{med}}|$ is first determined for each surrounding vector $\{\mathbf{U}_i | i = 1, \ldots, 8\}$. Next the median of these eight residuals, $r_{\text{med}}$, is determined and used to normalize the standard median test as follows:

$$\frac{|\mathbf{U}_{2D}(\text{med}) - \mathbf{U}_{2D}(i,j)|}{r_{\text{med}} + \epsilon_0} < \epsilon_{\text{thresh}}$$

The additional term $\epsilon_0$ is required to account for remaining fluctuations obtained from correlation analysis of otherwise quiescent or homogeneous flow. In practice this value should be set around $0.1 - 0.2$ pixel, corresponding to the mean noise level of PIV data [85] (see also section 5.5).

The efficiency of the normalized median test was demonstrated by Westerweel & Scarano [275] by applying it to a number of PIV experiments covering a wide range of Reynolds numbers. The probability density functions of both the standard and normalized median for these experiments is shown

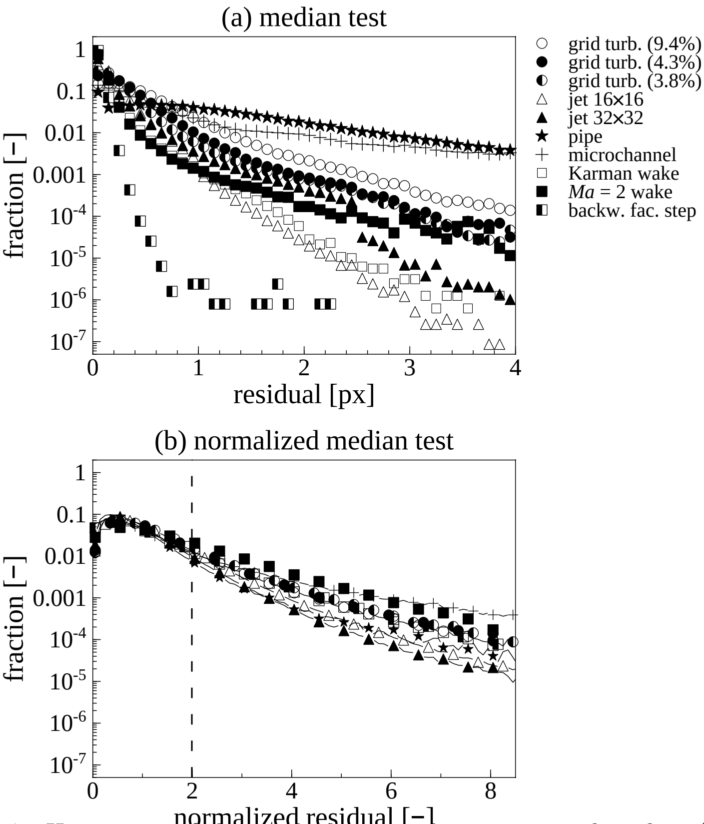


**Fig. 6.7.** Histograms of the residuals using the conventional median (a) and the normalized median (b) for a wide variety of experimental data as presented by Westerweel & Scarano [275].

in figure 6.7. Integration of the histograms of the residual for the normalized median indicated that the 90-percentile occurs for $r_{\mathrm{med}} \approx 2$. This meant that in all experiments investigated a single detection threshold labeled the largest 10% of residuals. The detection efficiency is less stringent for thresholds $r_{\mathrm{med}} > 2$ and vice versa.

The universality of this detection scheme makes it especially well suited for iterative PIV interrogation schemes such as those presented in section 5.4.4 and should be very suited for self-optimizing PIV algorithms.

### 6.1.6 Other Validation Filters

While fluid mechanical information can be used for validation, it is commonly only used indirectly by assuming that the investigated flow must observe a certain degree of continuity or coherency through the application of neighborhood operators. Another forms of data validation are possible through the use of redundant information that is available from time resolved, multi-frame PIV data [323] or from additional view points such as in stereo-PIV (see page 215).

The following describes a few more validation methods which are of lesser importance for various reasons. Their performance in comparison to the previous methods is outlined in table 6.1.

**Minimum correlation filter**

As mentioned earlier, a low correlation coefficient is indicative of a strong loss a particle match and may have a variety of causes. Thus, a validation filter may be very helpful in detecting problematic areas in the field of view. However it is of lesser importance for the actual validation of PIV data, as low correlation values do not necessarily point to invalid displacement readings.

**Peak height ratio filter**

In this case the correlation peak representative of the displacement reading is compared to the first noise peak in the correlation map. A low ratio of the peak heights may point to an inadequately seeded area and a higher likelihood that the measured displacement is questionable. In terms of validation it is less effective because mismatched areas may have high correlation coefficients especially when seeding levels are low.

**Signal-to-noise filter**

Here the signal-to-noise ratio in the correlation plane – defined as the quotient of correlation peak height with respect to the mean correlation level – is used to validate the data. However its use is questionable because mismatched particle images or stationary background features can also produce high levels of correlation.

**Table 6.1.** Various validation filters - outlier detection efficiency, number of required parameters and potential for self-optimization

| validation filter | no. of params. | detection efficiency | automated optimization | reference |
|---|---|---|---|---|
| magnitude | 1 | poor | simple | - |
| range | 2-4 | medium | simple | page 181 |
| dynamic mean | 2 | medium | difficult | page 183 |
| difference | 1 | high | possible | page 184 |
| median | 1 | high | possible | page 185 |
| normalized median | 1 | high | simple | page 185 |
| minimum correlation | 1 | poor | possible | page 187 |
| correlation peak ratio | 1 | poor | difficult | page 187 |
| correlation SNR | 1 | poor | simple | page 187 |
| reconstruction residuals[a] | 1 | high | simple | page 215 |

[a] only for stereo PIV data

### 6.1.7 Implementation of Data Validation Algorithms

Since there is no unique validation filter suitable for all applications, the common approach has been to apply a combination of several different filters in succession. By adjusting the validation parameters individually for each filter, high data validation rates can be achieved even if the individual filters are not operating at their optimum. This strategy is especially attractive when processing larger image quantities.

A successful validation procedure should be to collect as much a priori information about the flow field to be investigated as possible and to express this knowledge in the form of fluid mechanical or image processing operators. The first simple fluid mechanical operators have already been developed [88].