Performance of Inspection Systems

Factors Influencing Automated Defect Identification

The performance of automated inspection systems on the factory floor is a frequent source of controversy between equipment supplier and user – often caused by different perspectives. The overall performance as desired by the user does not depend on the technical performance of the inspection system alone. This paper intends to highlight different influence factors and to sensitize for a view of the complete process.

The proposition of any modern quality management is to have a stable and controlled production process instead of sorting out defective parts. Nevertheless, in practice process control has its technical limitations. Thus, automated inspection systems are often unavoidable at the end of the production line. There is, however, no such thing as a perfect inspection, neither human nor automated. This leads consequently to the question of system performance.

The original task of a final inspection system is to sort out bad parts. From the point of view of the QA, the slippage of bad parts still reaching the customer should be minimal. The intention of the production manager on the other hand is that the inspection should cause as few false rejects as possible. Both figures can be determined by taking samples from the good and the bad parts. Together with the fraction of correctly inspected parts, these figures can be summarized in a 2 x 2 contingency table (cf. fig. 1).

The final figures slippage and false-rejects are driven by the complete process, including the production process and the quality decision as well (see fig. 1): in production, defects of various types are generated with different frequencies, the inspection system does state these properties and depending on the actual quality specifications it is decided whether the part is 'ok' or 'not ok'.

The inspection system may fail, e.g. by making a wrong defect classification. The effect of such a failure depends on how often the particular defects appear (production statistics). Moreover, whether

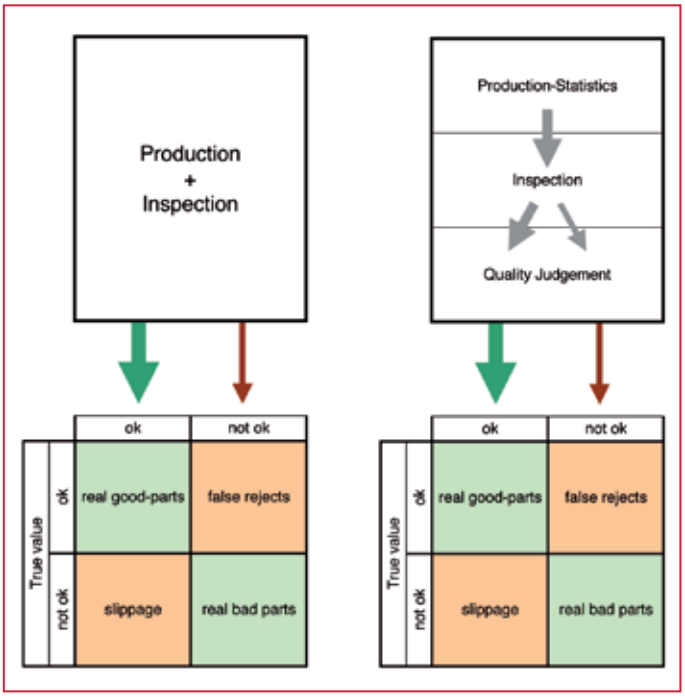

Fig. 1: From a naive point of view, the total production line including the inspection is a black-box, delivering correctly inspected parts (genuine good and genuine bad parts), and incorrect inspected parts (slippage and false rejects). A closer look however reveals that the inspection is just one link in the total process chain.

such a failure has any consequences depends on the subsequent quality judgment; maybe the misclassified defect still belongs to the same quality category.

Glass for example may contain small bubbles or inclusions, which are hard to distinguish (even for a human inspector) and are easily mixed up. However, this doesn't play a role when inclusions appear only very rarely (production statistics), or when bubbles and inclusions lead to the same product quality (quality judgment).

Inspection Systems

In the following, we will focus on inspection systems, which look for defects, e.g. bubbles, inclusion or scratches on glass products, or bumps and particles on coated sheets. Such tasks are in general performed by optical machine vision systems. However, the following reflections are also valid for any other imaging system, like ultrasound, x-ray etc, and may in parts also be translated to human inspection.

For physical reasons, there is no such thing as a 'perfect' inspection system, in the same manner as there are no perfect measurement systems. In any metrology device there are disturbances like noise, drift, nonlinearities etc. Even more, inspection system often cannot directly access the features that are originally relevant for the quality, but do record secondary quantities. It is for instance widely accepted that a particle's size is proportional to the light intensity that is scattered by the particle. These basic arguments show that no inspection equipment is immune against making errors.

Typically, defect inspection systems contain a signal processing chain of sensor + image pre-processing – detection – feature extraction – classification – quality judgment. First of all, a defect has to be found, i.e. to be detected, before it can be classified. The classification is based on the extraction of suitable image features. The quality judgment is based on the classification, which is mostly the type of defect. In addition to the defect classes, representing a categorical variable, an inspection system can also output continuous quantities. For example, the size of the defects is often used in the final quality decision.

One could assume that it should be possible to derive the performance of an inspection system by determining all measurement errors. In a procedure analogue to the GUM [1] one might then calculate the error propagation in the subsequent image processing and classification steps. Unfortunately, the algorithms involved are mostly by far too complex for such an approach. Moreover, the supplier would have to reveal his vital know-how. In practice it will therefore be necessary to carry out empirical tests, regarding the inspection system itself as a black box.

In such a test, the inspection result will be compared to some reference, e.g. to an offline analysis with a microscope. The reliability of any performance test, amongst other things concerning their reproducibility and repeatability, is completely dependent on the reliability of the reference. This needs to be questioned, particularly when we think about subjective decisions of human inspectors. For simplicity, we will in the following assume to have a reliable reference.

Inspection Failures

The failures of an inspection system have to be considered separately for categorical and

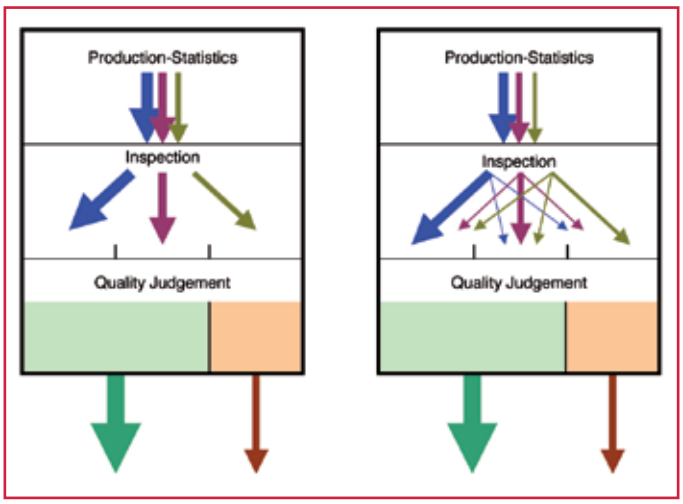

Fig. 2: Ideally, the inspection should unambiguously classify the defects according to their properties (l.h.s.). In reality, misclassifications are always possible (r.h.s.).

for continuous defect features. For categorical features, like the assignment of defect types, misclassifications can take place. For instance, a defect of type 'A' may be classified to be of type 'B' by the system (see fig. 2). The corresponding performance can be described in terms of classification rates p_{ij} (p_{BA}: real type 'A', inspection result type 'B'). These classification rates fulfill the condition $\Sigma_i\, p_{ij} = 1$.
The classification rates may depend on other parameters. In particular one may expect that classification errors occur more frequently for small defects, for which less information can be drawn from the image than for larger defects.

For any continuous measurand, systematic and random errors of measurement (repeatability precision with standard deviation σ) can be determined according to DIN 1319. Here, all random processes are supposed to be normally distributed (cf. fig. 3). Reproducibility, linearity and stability are additional relevant properties.

From the signal chain described above, one can take that detection failures also influence the performance. On the one side, there are undetected defects, immediately leading to slippage. This detection slippage will mostly depend on the defect size. On the other hand, the inspection system may detect a defect where there is in fact nothing wrong with the product. Such pseudo-defects can be caused by camera noise or by cosmic rays and may lead to false-rejects.

The picture of defect inspection systems drawn here is intentionally somewhat simplified in order to focus on the main aspects. Thus it is neither considered that there are systems with more than two quality categories, that parts may carry more than one single defect, or that systems exist that are trained instead of being parameterized.

Consequences of Inspection Failures

As an example, figure 4 demonstrates how classification errors map onto the overall performance. The defect types 'A' and 'B' are assumed to be harmless, while 'C' has to be rejected. The production statistics indicates how often the individual defect types occur: Multiplying the rates of the production statistics with the classification rates results in the weighted classification rates. Those are then summarized according to the quality categories into 'ok' and 'not ok'. As a result,

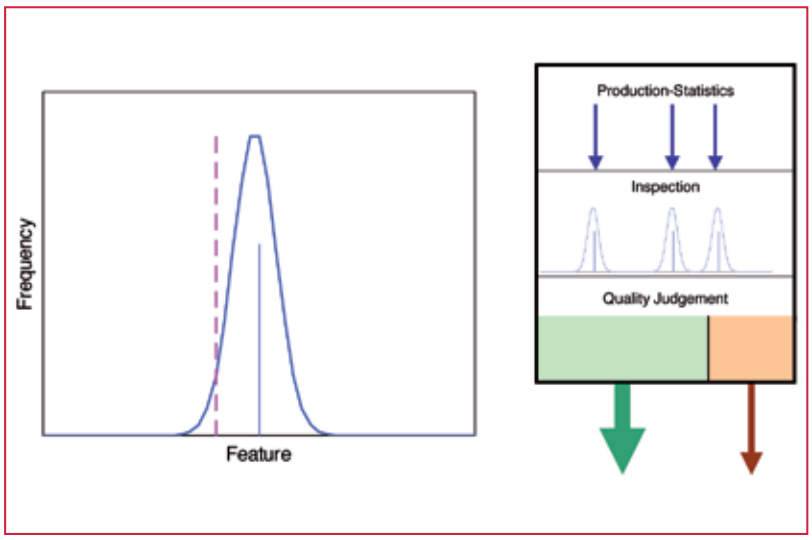

Fig. 3: The measurement value typically deviates from the true value (vertical line) due to a bias (difference to average of many measurements) und due to random contributions (bell-shaped curve). Consequently, measured values can sometimes lie on the other side of the tolerance limit (dashed vertical line), leading to slippage or false-rejects.

we directly arrive at slippage- and false-reject rates. This consideration refers to parts carrying defects only; for a practical application defect-free parts have to be taken into account as well.

For an ideal inspection system, all values on the diagonal of the classification rate matrix would be $p_{ii} = 100\,\%$, while all off-diagonal values would equal 0. From the diagram in figure 4 one can read that in such a case, slippage and false-reject rates would vanish.

In many cases, continuous defect features are subject to a single-sided limit; in particular the defect size may often not exceed an upper tolerance limit. The closer the measured value comes to this limit, the larger are the effects of measurement errors, more frequently leading to mistakes in the final quality decision. In the extreme case, when a defect is exactly on the tolerance limit, even a perfect inspection would always state the defect to be 'ok' in 50 % of the cases and to be 'not ok' in the other 50 %.

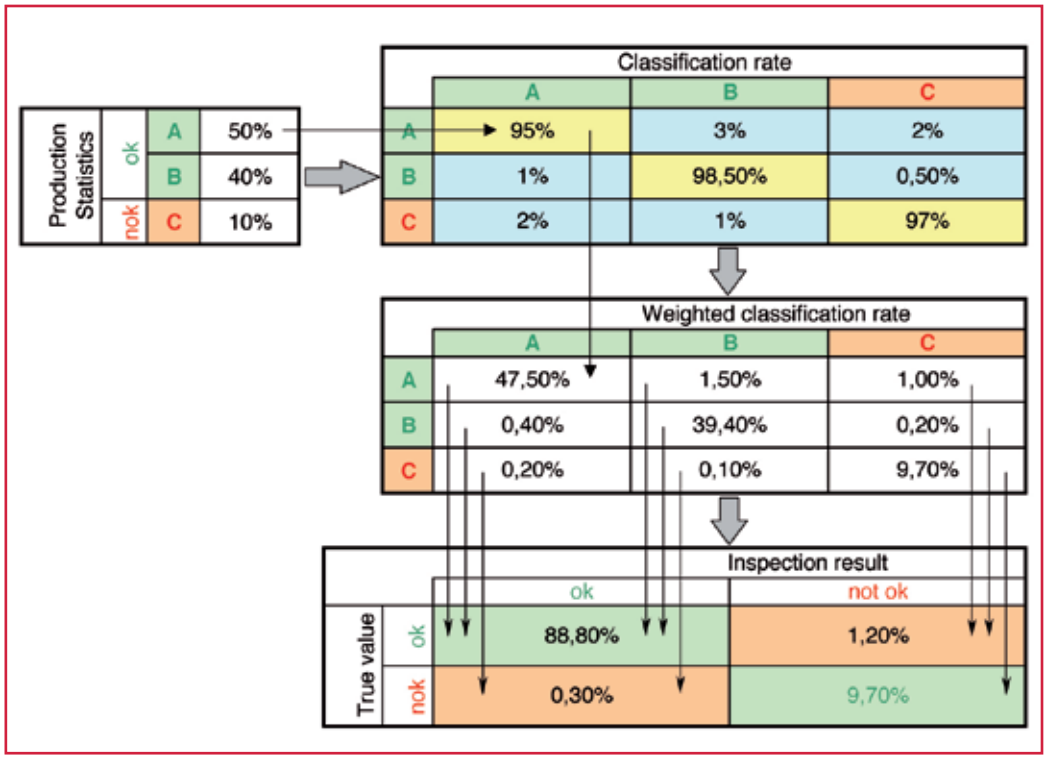

Fig. 4: Consequences of classification errors: the defect's frequencies (production statistics) and the classification rates p_{ij} are taken to derive the weighted classification rates, i.e. the statistics of the inspection results. These are summarized to the two quality categories. In comparison to the true defect types, slippage- and false-reject rates result.

The said arguments apply to a single measurand, but in general all possible values have to be taken into account. Mathematically speaking, the production statistics has to be convolved with the distribution of the measurement error (cf. fig. 5).

Undetected defects (detection slippage) are neither classified nor measured in any way by the inspection system and end up directly at the customer. The resulting slippage corresponds to the product of the detection-slippage rate and the ratio of 'not ok'-defects in the production statistics. Correspondingly, the contribution of pseudo-defects to false-rejects depends on how often their features statistically lead to an 'ok' or 'not ok' quality decision.

Validated Inspection Systems?

In Germany, an established qualification procedure for measuring instruments is the so-called 'Messmittelfähigkeit' (measurement system and equipment capability). Users of inspection equipment often ask why their equipment couldn't simply be qualified in a similar way. A closer look at the procedure reveals the problems: 'Messmittelfähigkeit' refers to a single continuous measurand. The measurement error is related to a double-sided tolerance band to assure a small false reject rate. The procedure evaluates the error by the measurement instrument only and includes the quality criteria by including the tolerance limits. Using a calibrated standard assures that the true value of the measurand equals the nominal value in the centre of the tolerance band [2].

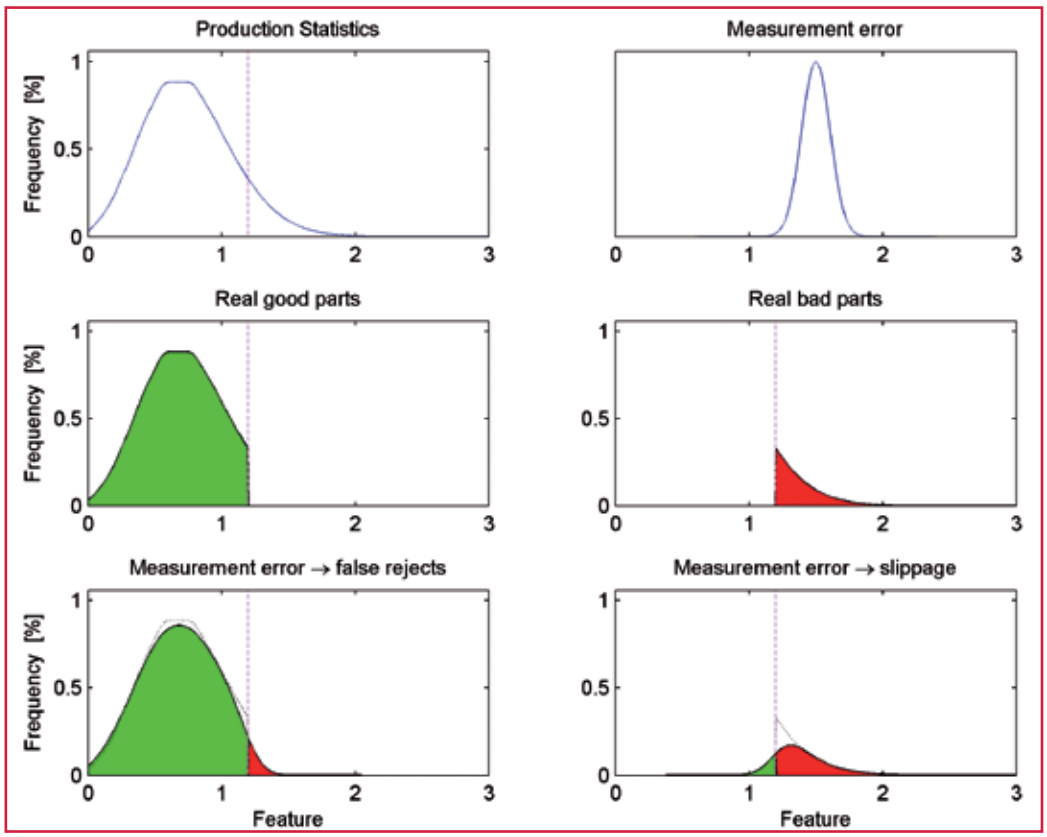

Fig. 5: Impact of the measurement error: the production statistics is shown on the upper l.h.s., the upper r.h.s graph indicates the error distribution. The production statistics is split into the o.k. and the n.o.k. parts according to the tolerance limit (middle row). The washing-out of these distributions due to the measurement error causes fractions to cross the tolerance limit, leading to slippage and false-rejects (lower row).

A stable and well-controlled manufacturing process should yield parts for which the measured value can be expected to comply with the nominal value. Under this condition the 'production statistics' corresponds to the supposition underlying the described method. The manufacturing process, however, is not necessarily stable: false rejects and slippage can grow very large as soon as the manufacturing gets out of control, leading to parts close to one of the tolerance limits. Hence, a measurement instrument complying with the described standard is by no means any warranty against false-rejects or against slippage – which however often is the implicit expectation of many users.

Quantities describing defect features do in general not posses a single fixed expectation value. Hence, without knowledge about the feature distribution, i.e. the production statistics, inspection systems cannot be validated in a meaningful manner. Moreover, the method for 'Messmittelfähigkeit' has to be transferred to categorical quantities. Finally, in the present context continuous feature quantities are mostly not subject to double-sided, but to single-sided tolerance limits. The described method cannot be applied easily to such single-sided limits.

Conclusion

The overall performance in the sense of false-rejects and slippage depends equally on the production statistics, on the technical performance of the inspection equipment, and on the actual quality criteria. Therefore, inspection equipment cannot be designed, set up or validated by itself, independent from the complete process chain. From this interdependency it follows also that any change in the production process or any modification of the quality specifications immediately influences the overall inspection performance.

In most cases inspection systems are individually adapted to the particular application. Still, the steps detection, classification and measurement of continuous quantities are contained in the majority of the systems. As shown here, the respective system deficiencies can precisely be defined and can be used to derive the consequences for the overall performance in a quantitative way. This can be used as building blocks of an individual inspection system validation.

References

[1] Guide to the Expression of Uncertainty in Measurements', DIN V ENV 13005

[2] E. Dietrich, A. Schulze, S. Conrad: Eignungsnachweis von Messsystemen, Hanser Verlag, 2008

▶ **Author**
Dr. Ralph Neubecker, Project Manager Optical Metrology

▶ **Contact**
Schott AG, Mainz, Germany
Tel.: +49 6131 66 0
Fax: +49 6131 66 2000
ralph.neubecker@schott.com
www.schott.com