Towards a Reliable Statistical Oracle and its Applications

Johannes Mayer

mayer@mathematik.uni-ulm.de

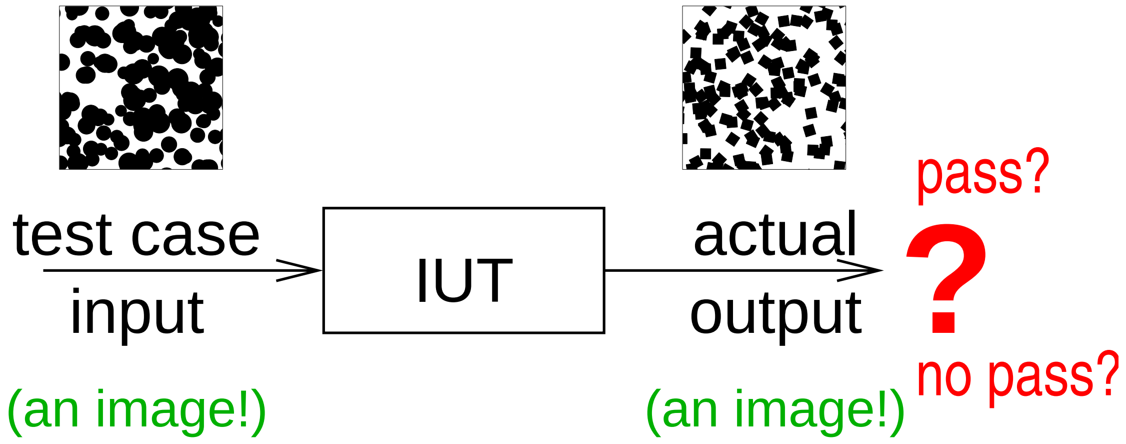
Abt. Angewandte Informationsverarbeitung
Universität Ulm

Overview

1. Motivation: How to test image processing applications?
2. What is an oracle?
3. Typical examples of oracles
4. What about random testing?
5. A pattern for the Statistical Oracle
6. Necessary image processing basics
7. How to apply the Statistical Oracle to test image processing applications?
8. Conclusion

Motivation

How to test an image processing application?

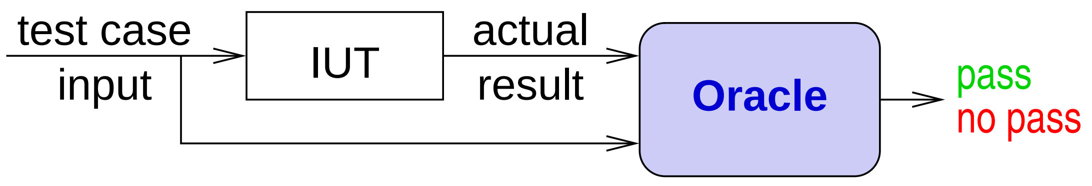


- Complex input and output
- How to determine the expected output?
- Also: How to generate many complex inputs?

Oracles

What is an oracle?

Overview:



An **oracle** is necessary to decide whether a test case passes or not.

Oracles
Examples of Oracles

- Judging:
  Tester evaluates pass/no pass
- Prespecification:
  Solved examples, approximation, parametric, . . .
- Gold Standard:
  Trusted system, voting, regression, . . .
- Organic:
  Reversing, built-in test, . . .

$\rightsquigarrow$ all oracles only applicable in special cases

Random Testing

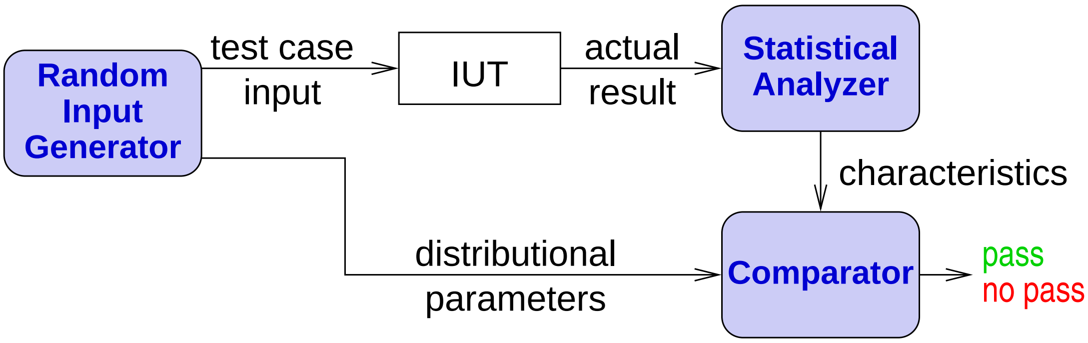
Random Testing is testing using randomly generated inputs.

Complex inputs (e. g. images) are not easy to specify
$\rightsquigarrow$ random generation would pay off

But what about the outputs?
$\rightsquigarrow$ An appropriate oracle is needed for automation

Statistical Oracle
Pattern

Overview:



Necessary:

Mean, Variance, and so on of characteristics must be known (explicit formula or algorithm)

Statistical Oracle
Necessary Statistical Basics (1)

Notation:

- Let $X_1, \ldots, X_n$ denote i. i. d. random variables (the inputs of the comparator – X_i belongs to the i th test case)
- The mean μ and variance σ^2 of X_i is unknown.
- The sample mean of these random variables $X_1, \ldots, X_n$ is

$$\overline{X_n} := \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Notation (continued):

- According to the central limit theorem, it holds that

$$\frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$$

  for $n \to \infty$.
- Thus, $\overline{X_n}$ can be regarded as approximately normally distributed with mean μ and variance σ^2/n if $n \geq 30$ (a common rule of thumb)
- The greater n gets, the less likely deviations from μ become

Notation (continued):

- The sample variance

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X_n})^2$$

  of the random variables $X_1, \ldots, X_n$ (approaches σ^2 as n goes to infinity)
- x_i, $\overline{x_n}, s_n^2$ denote realizations
- μ_0 denotes the expected mean of X_i

Statistical Oracle
Statistical Decision Test

Given the (maximum) probability of $\alpha \in (0, \frac{1}{2})$ of a Type I error (i. e. α can be controlled)

Type I error: reject H_0 under H_0
Type II error: H_0 not rejected under H_1

Construction:
Intersection-union test with two one-sided t-tests

$H_0 : \mu \notin [\mu_0 - \Delta, \mu_0 + \Delta]$ (buggy)
$H_1 : \mu = \mu_0$ (correct)

($\Delta > 0$ can be chosen arbitrarily)

Type I error: false rejection of H_0 (i. e. a buggy IUT passes)

Statistical Oracle
Statistical Decision Test: Decision Rule

Reject H_0, i. e. the implementation passes, if

$$\frac{\overline{x_n} - (\mu_0 - \Delta)}{s_n/\sqrt{n}} \geq t_{n-1,\alpha}$$

and

$$\frac{\overline{x_n} - (\mu_0 + \Delta)}{s_n/\sqrt{n}} \leq -t_{n-1,\alpha}$$

where $t_{n-1,\alpha}$ denotes the $(1 - \alpha)$-quantile of the t-distribution with $n - 1$ degrees of freedom

Statistical Oracle
Statistical Decision Test: Type II Error

- Probability α of Type I error already controlled
- Probability β of Type II error:

$$\begin{aligned} & \mathbb{P}(\text{decision for } H_0 \mid H_1) \\ = \; & \mathbb{P}\left(t < t_{n-1,\alpha} - \frac{\Delta}{s_n/\sqrt{n}} \vee t > -t_{n-1,\alpha} + \frac{\Delta}{s_n/\sqrt{n}} \mid \mu = \mu_0\right) \end{aligned}$$

  with $t = \frac{\overline{x_n} - \mu_0}{s_n/\sqrt{n}}$

- Consequences:
  - Necessary: $\frac{\Delta}{s_n/\sqrt{n}} > t_{n-1,\alpha}$
  - β decreases, when
    - α, Δ, or n increases
    - σ decreases (and consequently s_n decreases)

Statistical Oracle
Statistical Decision Test: Type II Error (2)

Solution:

- For $\alpha = 1/4$ determine n such that $\beta \leq 1/4$ (Δ and σ are given)
  ($\rightsquigarrow$ numerically!)
- **BPP** algorithm: each answer is correct with probability at least $3/4$
- Repeat this m times and make a majority decision to achieve arbitrariliy small error probabilites

Statistical Oracle
Consequences

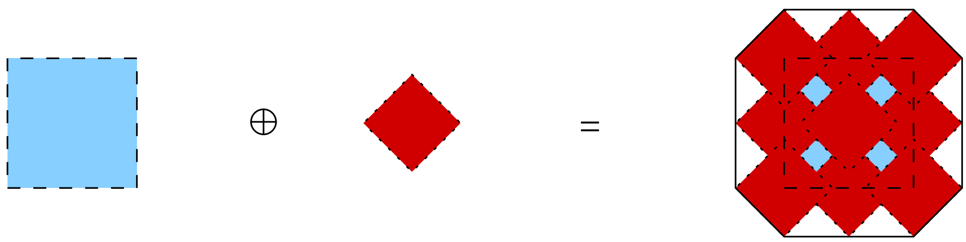
- Cannot decide for a single test case, but for a set of test cases
- Decision is not always correct, but with arbitrarily high probability
- Does not check the whole result, but only some properties
  ($\leadsto$ only correctness with respect to the checked property and with respect to the used inputs)

Image Processing
Premliminaries: Minkowski addition and dilation

The Minkowski addition $A \oplus B$ is defined as

$$A \oplus B := \{x + y : x \in A, y \in B\}.$$

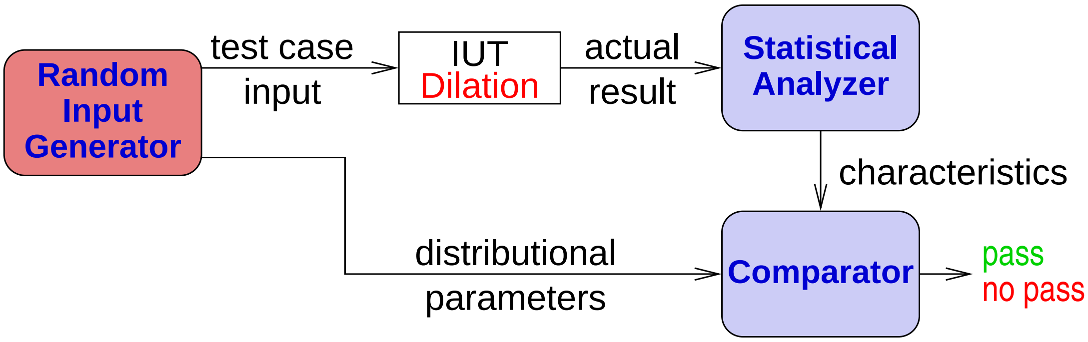
The set B is in this case called the structuring element.



A B $A \oplus B$

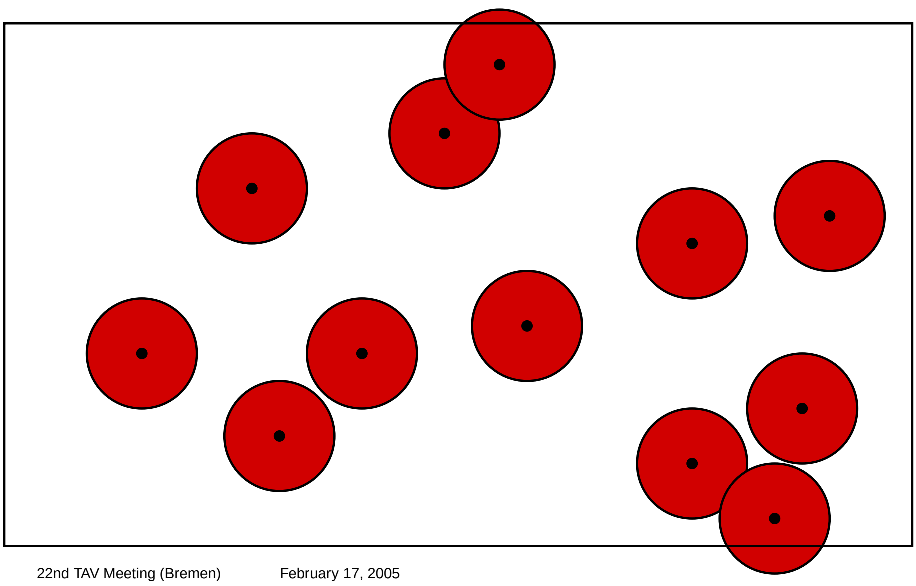
$\delta_B(A)$ denotes the dilation of A with the structuring element B and is defined as $\delta_B(A) := A \oplus \check{B}$

Testing of Dilation
Random Input Generator: Overview



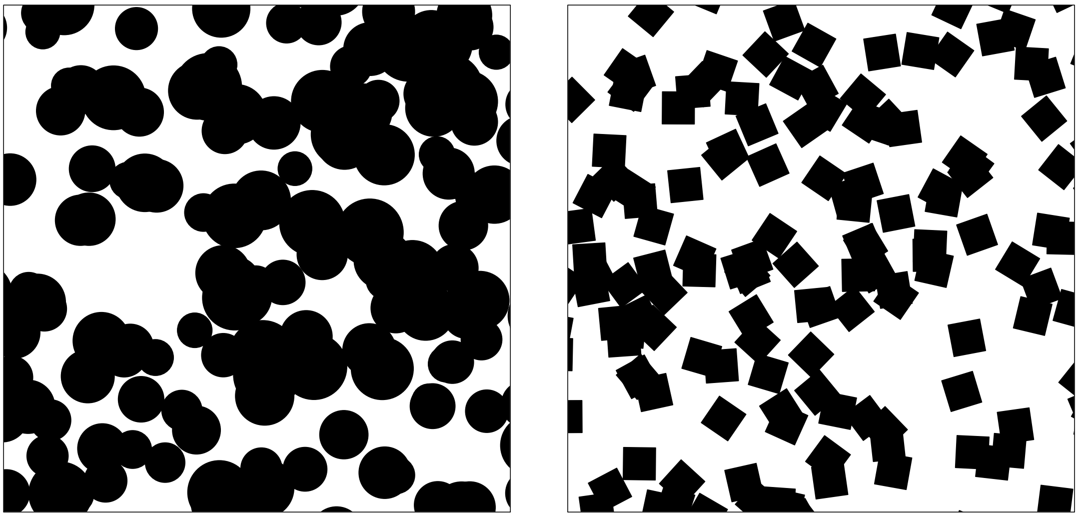
Testing of Dilation
Random Input Generator: Boolean Model

The Boolean model is simply the union $\bigcup_i (B_i + X_i)$ of i. i. d. random grains B_i each translated into another point X_i of the underlying Poisson process.



Testing of Dilation
Boolean Model: Examples

Examples:



Testing of Dilation
Boolean Model: Useful Property

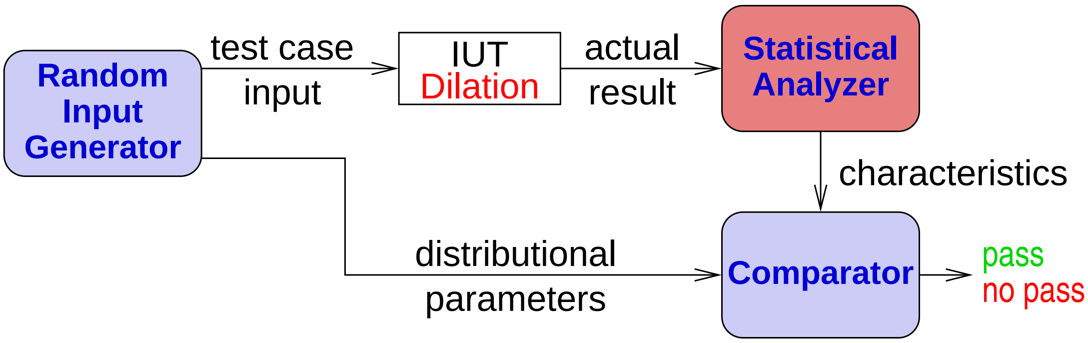
What happens, when the Boolean model $\bigcup_i (B_i + X_i)$ is dilated by B?

$$\delta_B\left(\bigcup_i (B_i + X_i)\right) = \left(\bigcup_i B_i + X_i\right) \oplus \check{B}$$
$$= \bigcup_i ((B_i + X_i) \oplus \check{B}) = \bigcup_i (B_i \oplus \check{B}) + X_i,$$

The result is the Boolean model with grains $B_i \oplus \check{B}$.

For example, if B_i is a disc with random radius R_i and B is a disc with radius r, $B_i \oplus \check{B}$ is simply a disc with random radius $R_i + r$.

Testing of Dilation
Characteristics: Overview



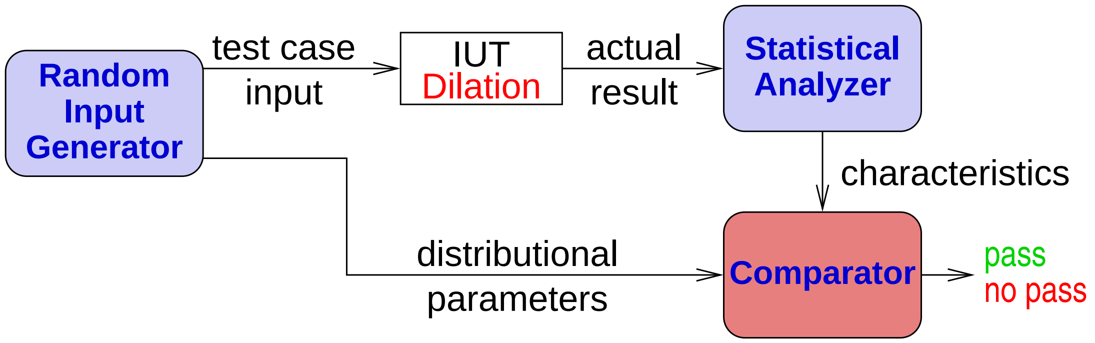
Testing of Dilation
Characteristics: Formulae

Choosing specific area A_A, specific boundary length L_A, and specific Euler number χ_A
$\rightsquigarrow$ explicit formulae are known:

$$
\begin{aligned}
A_A &= 1 - \exp(-\lambda \overline{A}) \\
L_A &= \lambda \overline{L} \exp(-\lambda \overline{A}) \\
\chi_A &= \lambda \left(1 - \frac{\lambda \overline{L}^2}{4\pi}\right) \exp(-\lambda \overline{A})
\end{aligned}
$$

- $\overline{A}$ mean area of the primary grain
- $\overline{L}$ mean boundary length of the primary grain
- λ intensity of the Poisson process

Testing of Dilation
Comparator



The comparison is done with the presented intersection-union test

Testing of Dilation
Putting the Pieces together

1. The Random Test Input Generator generates realizations of the Boolean model. Passes the characteristics of the Poisson process (λ) and of the primary grain ($\overline{A}$ and $\overline{L}$) to the Comparator.
2. The Statistical Analyzer computes the estimators for A_A, L_A and χ_A. Each such estimator can be modeled as a random variable X_i. Then A_A, L_A and χ_A are the expected means, respectively, of these random variables (i. e. μ_0).
3. The Comparator accumulates the realizations x_i of each such random variable X_i for n outputs and computes the realization $\overline{x_i}$ of the sample mean. Finally, it decides using an intersection-union test — as described — whether the IUT passes or not. (The error probability α can also be given to the Comparator.)

Conclusion

- Statistical Oracle for random testing if mean etc. is known
- Based on an intersection-union decision test
- Error probabilities can be chosen as needed through repetition
- Random testing of image processing applications pays off, due to complex inputs
- Boolean model can be used to test dilation
- Relevant characteristics: area, boundary length, Euler number

Perspectives

- Duality between dilation and erosion resp. Minkowski addition and Minkowski subtraction
  $\rightsquigarrow$ Erosion can also be tested
- Distance transform usually used to compute dilation and erosion
  $\rightsquigarrow$ Distance transform can be tested indirectly

References

- Robert V. Binder: *Testing Object-Oriented Systems.* Addison-Wesley, 1999.
- Johannes Mayer, Ralph Guderlei: *Test Oracles Using Statistical Methods.* In: Lecture Notes in Informatics P-58, Gesellschaft für Informatik, Köllen Druck+Verlag GmbH, 2004, pp. 179–189.
- Johannes Mayer: *On Testing Image Processing Applications With Statistical Methods.* Preprint, 2005. (accepted by Software Engineering 2005)