DETECTION OF FAULTY GLUCOSE MEASUREMENTS USING TEXTURE ANALYSIS

Nevine Demitri and Abdelhak M. Zoubir

Signal Processing Group, Institute of Telecommunications
Technische Universität Darmstadt, Merckstr. 25, 64283
Darmstadt, Germany
{ndemitri, zoubir}@spg.tu-darmstadt.de

ABSTRACT

Faults occurring in hand-held blood glucose measurements can be critical to patient self-monitoring, as they can lead to unnecessary changes of treatment. We propose a method to detect faulty glucose measurement frames in devices that use a camera to estimate the glucose concentration. We assert that texture, as opposed to intensity, is able to differentiate between correct and false glucose measurements, regardless of the given blood sample. The co-occurrence based textural features energy, maximum probability and correlation prove to be suitable for our detection application. We calculate kinetic feature curves and use a hypothesis testing approach to detect faulty measurements. Our method is able to detect a faulty measurement after less than one third of the time, which would usually be needed. The validation of our method is done using a real data set of blood glucose measurements obtained using different glucose concentrations and containing both correct and faulty measurements.

Index Terms— GLCM-based features, texture analysis, anomaly detection, blood glucose measurement

1. INTRODUCTION

Regular self-monitoring using small hand-held devices is of utmost importance for diabetes patients. In [1], a novel, invasive, photometric glucose measurement system is introduced, which uses a much smaller blood sample than state-of-the-art devices to minimize the induced pain. Here, the blood sample extracted from the patient's finger is placed on a chemical test strip. The temporal reaction between the chemical and the blood glucose induces a color change of the test strip that is observed optically using a camera. Using methods of image segmentation, the region of interest (ROI) is identified [2, 3] and the final convergence value of the intensity is estimated and mapped to the underlying glucose concentration.
In some cases, the blood drop can be insufficient or badly placed on the test strip, such that the chemical reaction does not take place correctly. Also, artifacts such as large dust particles or air bubbles can disturb the surface leading to an insufficient size of the ROI. Such cases have to be detected

by the device and the measurement has to be broken off and repeated, as to not deliver a false measurement value to the patient. We propose to use methods of texture analysis to detect such cases. It is intuitive to analyze the images in our application w.r.t. texture as we are dealing with different physical textures: (i) the dry chemical test strip, (ii) the wet test strip when the blood sample is applied, (iii) the presence of artifacts or air bubbles, (iv) as well as different textures for different stages of the chemical reaction and different underlying glucose concentrations.

In general, texture describes the spatial and statistical distribution of gray tones in a frame. Many methods exist to describe texture; they can be divided into four categories according to [4]: statistical, geometrical, model-based and signal processing. The most common method was proposed by Haralick in [5]. It assumes that the texture information in an image is contained in the average spatial relationship different gray-levels have to one another. The so-called gray-level co-occurrence matrix is defined to describe this and statistical features are derived from it. Haralick's texture analysis method has been used for a wide range of applications such as biomedical and aerial image segmentation as well as quality control, to describe different regions of an image or detect anomalies [5–7].

Our contributions lie in proposing an algorithm that uses texture features to analyze faulty glucose measurements. We identify features that are suitable for anomaly detection in our application and use a hypothesis testing framework to perform the detection. We show that using our algorithm, faults are detected early, hereby avoiding giving a false measurement to the patient and lowering unnecessary energy consumption of the device.

The remainder of the paper is organized as follows: First, the underlying photometric measurement principle is described in more detail in Section 2. Section 3 presents the texture analysis methods used in this work. Section 4 illustrates the proposed algorithm, followed by the description of the data set and the results in Section 5. Finally, a conclusion is given in Section 6.

2. PHOTOMETRIC MEASUREMENT PRINCIPLE

The photometric measurement principle, illustrated in Fig. 1, is a common principle to measure the concentration of an analyte in a fluid, such as the concentration of glucose in a blood sample. It relies on a chemical reaction between the blood glucose and a chemical agent on a test strip that yields a change of color [8]. The test strip is illuminated and
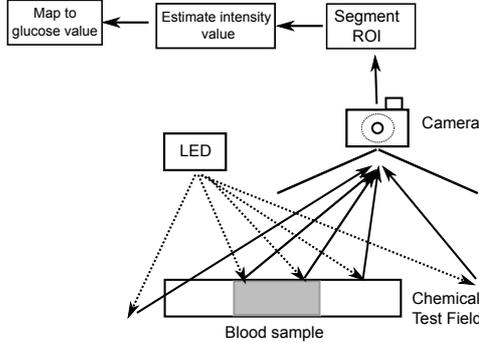
Fig. 1: Underlying photometric measurement principle.

the reflectance behavior is observed using a camera. Using methods of image segmentation [2, 3], the region of interest, where the chemical reaction is taking place is found. The corresponding amount of reflected light at the current frame is estimated. Finally, the chemical reaction converges and the final amount of reflected light is directly related to the underlying glucose concentration value.
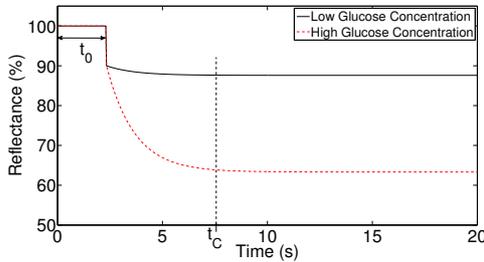
Fig. 2: An idealized model of the kinetic curve of the chemical reaction for high and low glucose concentrations.

The temporal development of the chemical reaction, termed kinetic curve, can be typically subdivided into three stages as depicted in Fig. 2.

1. Constant intensity stage at $t < t_0$: the reaction between the glucose and the chemical agent has not started.

2. The moistening period at $t = t_0$: the blood drop is detected by the chemical and the reaction starts; this stage is characterized by a rapid drop of the intensity value.

3. The convergence stage: this stage exhibits a slow decrease of the intensity and finally convergence at $t = t_C$, where

t_C depends on the underlying glucose concentration and usually lies between $5 - 10s$.

Fig. 3 depicts examples of the observations, obtained by the camera, at the different time stages of the chemical reaction and for different underlying glucose concentrations. Fig. 3a shows an observation at time $t < t_0$, where the reaction has not yet taken place. Fig. 3b, 3c show observations after convergence ($t > t_C$) for low and high glucose values, respectively. Errors may occur in the measurement process that do not lead to the typical observations, hereby leading to false estimates of the underlying glucose values. Fig. 3d -3f exemplify such cases obtained for $t > t_C$, where in Fig. 3d and 3e the blood drop is either insufficient or incorrectly placed on the test strip. In Fig. 3f the test strip is not placed correctly in the observation area of the camera, leading to an occlusion of the ROI. When analyzing the given frames, we observe that the different cases are not only uniquely marked by their intensity behaviors but also by their texture. The advantage of texture is that it shows unique characteristics for faulty measurements irrespective of the underlying glucose concentration. This is opposed to intensity that is closely related to the underlying glucose concentration. We therefore use the texture information in the frames to identify erroneous measurements. We show that using our method we can detect faulty measurements at the beginning of the moistening period at $t = t_0$, instead of having to wait for the final convergence value at $t = t_C$.
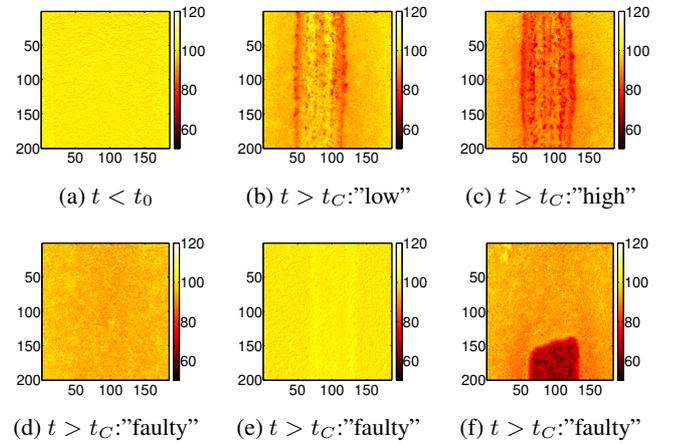
Fig. 3: Examples of the chemical reaction observed by the camera. 3a shows an example at $t < t_0$, 3b "low" and 3c "high" show observations of $t > t_C$ for low and high glucose values, respectively. 3d-3f show examples of faulty measurements for $t > t_C$.

3. GLCM-BASED TEXTURE ANALYSIS METHODS

In [5], Haralick proposed a gray-level co-occurrence texture method, that has become very popular for texture segmentation applications. Given an image I of size $N_x \times N_y$ with the

gray-levels $G = \{1, 2, ..., N_G\}$ the method analyzes the spatial relationships of the occurring gray-levels. The underlying assumption made is that texture information can be specified using a square matrix of relative frequencies with which two neighboring pixels with defined gray-level values (i, j) occur at a certain distance d and angle α. This is described using the so-called gray-level co-occurrence matrix (GLCM) $P(i, j; d, \boldsymbol{\alpha}, L)$. 8-pixel neighborhoods are considered, such that the angles used are $\boldsymbol{\alpha} = \{0°, 45°, 90°, 135°\}$ as well as their symmetric counterparts. The size of the matrix L is typically set to the number of occurring gray-levels N_G, however, coarser quantization can be used to accelerate computation and reduce noise [9]. After calculating the GLCM, statistical features can be calculated using the co-occurring relative frequencies, i.e. the entries of the GLCM. The features can be defined to be directional calculating the GLCM for one angle α or rotationally invariant by calculating the GLCMs of all angles and averaging over the resulting features [5].

For our application we need features that can distinguish between a correct measurement and a false measurement, regardless of the underlying glucose value. We analyzed all GLCM-based features proposed in [5] and [7], and identified three features that proved to be sensitive to faulty measurements: the energy (E), the maximum probability (M) and the correlation (C). They are given as follows:

$$s^E = \sum_{i=1}^{G} \sum_{j=1}^{G} P(i, j; d, \boldsymbol{\alpha}, L) \quad (1)$$

$$s^M = \max_{ij} P(i, j; d, \boldsymbol{\alpha}, L) \quad (2)$$

$$s^C = \frac{\sum_{i=1}^{G} \sum_{j=1}^{G} ij(P(i, j; d, \boldsymbol{\alpha}, L) - \mu_x \mu_y)}{\sigma_x \sigma_y}, \quad (3)$$

where $\mu_x, \mu_y, \sigma_x, \sigma_y$ are the means and standard deviations of the marginal distributions of $P(i, j; d, \boldsymbol{\alpha}, L)$. For cases such as in Fig. 3d - Fig. 3f the GLCM is characterized by a very high peak at the first couple of entries of the diagonal as can be seen in Fig. 4a. Contrarily, for cases such as Fig. 3b and Fig. 3c the GLCM shows values spread over the diagonal as can be seen in Fig. 4b. This behavior is observed regardless of the underlying glucose value. The features s^E, s^M, s^C measure this behavior and are therefore suitable for identifying incorrect measurements.

4. PROPOSED ALGORITHM

The aim of this work is to identify faulty blood glucose measurements in an early stage of the chemical reaction. Once a measurement is identified as being faulty, the measurement procedure is broken off and the patient is asked to repeat the measurement, thus avoiding the estimation of a false blood glucose concentration. Detecting the false measurement at an early stage also saves unnecessary power consumption and time. The proposed algorithm is summarized in Fig. 5 and is
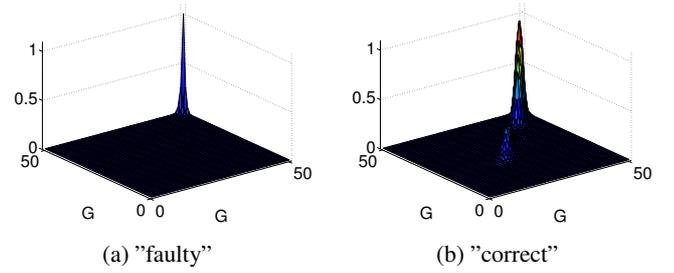
(a) "faulty"　　　(b) "correct"

Fig. 4: Examples GLCMs for a faulty (4a) and correct (4b) measurement.
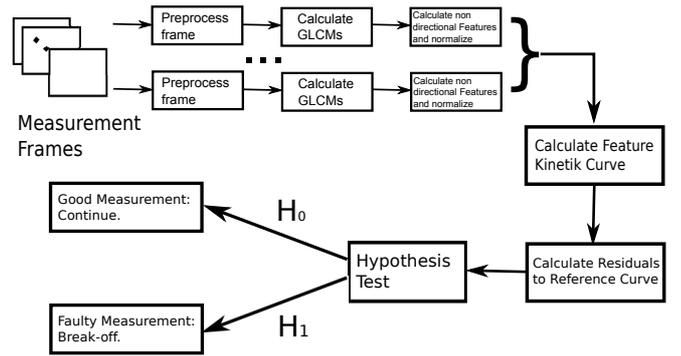
Fig. 5: Flow graph of the proposed algorithm for detection of faulty glucose measurements.

presented subsequently

Step 1: First the observed frame is preprocessed as described in [2]. This involves the normalization of the frames using a set of calibration frames recorded at the beginning of every measurement. Furthermore, the images are cropped to the relevant area and binned to reduce noise.

Step 2: Afterwards, the GLCMs $P(i, j; d, \boldsymbol{\alpha}, L)$ are calculated for each frame. Our analysis shows that using $L = N_G$ produces the same accuracy of results as $L = N_G/2$. We use $L = N_G/2$ to accelerate computation. The GLCMs are calculated using a distance of $d = 1$ as we are dealing with micro-structured texture. We calculate $N_{\boldsymbol{\alpha}} = 4$ GLCMs for each frame, corresponding to the $N_{\boldsymbol{\alpha}} = 4$ angles $\boldsymbol{\alpha} = [0°, 45°, 90°, 135°]$.

Step 3: Then, the features are calculated for each GLCM, resulting in a feature vector

$$\mathbf{s}^i = [s^i_{\boldsymbol{\alpha}(1)}, s^i_{\boldsymbol{\alpha}(2)}, s^i_{\boldsymbol{\alpha}(3)}, s^i_{\boldsymbol{\alpha}(4)}], \quad i \in \{E, M, C\}$$

for each frame. To make the features rotationally invariant the mean of the features w.r.t the angles is calculated, hereby, building following feature vector for each frame

$$\mathbf{s} = \begin{bmatrix} \frac{1}{N_{\boldsymbol{\alpha}}} \sum_{k=1}^{N_{\boldsymbol{\alpha}}} s^E_{\boldsymbol{\alpha}(k)} \\ \frac{1}{N_{\boldsymbol{\alpha}}} \sum_{k=1}^{N_{\boldsymbol{\alpha}}} s^M_{\boldsymbol{\alpha}(k)} \\ \frac{1}{N_{\boldsymbol{\alpha}}} \sum_{k=1}^{N_{\boldsymbol{\alpha}}} s^C_{\boldsymbol{\alpha}(k)} \end{bmatrix}. \quad (4)$$

The feature vector **s** is normalized using linear scaling to unit range as described in [10]

$$\tilde{\mathbf{s}} = (\mathbf{s} - \mathbf{l}) \oslash (\mathbf{u} - \mathbf{l}),$$

where $\mathbf{l} = [l^E, l^M, l^C]^T$ and $\mathbf{u} = [u^E, u^M, u^C]^T$ contain the lower and upper values of the features (E), (M) and (C) and $\oslash$ denotes element-wise division of the vectors.

Step 4: Steps 1-3 are performed for every frame in a measurement, producing a feature kinetic curve $k^i(n)$ with $i \in \{E, M, C\}, n = 1, ..., N$, where N is the total number of frames.

Step 5: The residuals of $k^i(n)$ and a reference kinetic curve for the features, $k_{\text{ref}}^i(n)$ are calculated

$$r^i(n) = k^i(n) - k_{\text{ref}}^i(n). \tag{5}$$

The reference kinetic curve is obtained using the real data. This is done by calculating the feature kinetic curves of all the correct measurements and take the median of all curves to be the reference curve

$$k_{\text{ref}}^i(n) = \underset{m \in M_C^i}{\text{median}}\{k_m^i(n)\},$$

where M_C^i is a set containing all feature kinetic curves of correct measurements of the corresponding features.

We assume the residuals $r^i(n)$ to be independently and identically distributed (i.i.d.). They can be approximated by a random process R^i drawn from a Gaussian distribution according to $f_{R^i} \sim \mathcal{N}(0, (\sigma_1^i)^2)$, where σ_1^i is unknown and has to be estimated using the given data. For the case of faulty measurements the residuals follow a Gaussian distribution according to $f_{R^i} \sim \mathcal{N}(\mu^i, (\sigma_2^i)^2)$, where $\mu^i > 0$ and σ_2^i is unknown. Fig 6 shows an example of the distribution of residuals for correct measurements and faulty measurements for the feature (M).

Step 6: A hypothesis test is performed, where

$$\mathcal{H}_0 : f_{R^i} \sim \mathcal{N}(0, (\sigma_1^i)^2), \quad \text{correct measurement} \tag{6}$$
$$\mathcal{H}_1 : f_{R^i} \sim \mathcal{N}(\mu^i, (\sigma_2^i)^2), \quad \text{faulty measurement.}$$

Here, $\mu^i > 0$, σ_1^i and σ_2^i are both unknown. Given the i.i.d. assumption, we use a Neyman-Pearson approach for the hypothesis testing that results in the following test

$$T(r^i(n)) = \frac{1}{N} \sum_{n=1}^{N} r^i(n) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda^i, \tag{7}$$

where λ^i is determined using a nominal value for the probability of false alarm $P_{\text{FA}} = \int_{\lambda^i}^{\infty} P(r^i(n)|\mathcal{H}_0)$.

If the null hypothesis is rejected, we assume the measurement to be faulty. The measurement procedure is broken off and the user is given an error message and asked to repeat the measurement. As long as we accept the null hypothesis, we keep on performing the test for each new incoming sample,
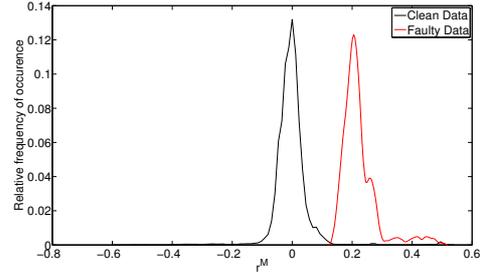
Fig. 6: Relative frequency of occurrence of the residuals of clean and faulty measurements.

as we want to cover the cases where the fault happens during the measurement. This can occur for example, when due to a strong movement of the device, the chemical test strip shifts and is not under the observation area of the camera.

5. SIMULATIONS AND RESULTS

5.1. Data Set

We use a real data set consisting of $N_T = 162$ measurements that was recorded using a prototype of the device presented in [1]. The measurements were made using whole blood samples with $N_G = 16$ different underlying glucose values. Each measurement consists of $N_f = 582$ frames taken with a sampling rate of 30 fps. The first $N_C = 32$ frames are used for calibration purposes; the chemical reaction can be observed in the remaining $N_R = 550$ frames. The measurements are labeled such that for each measurement we have: (i) information on the underlying glucose concentration used to perform the measurement, (ii) the estimated glucose concentration value estimated by the device and (iii) the convergence time needed to estimate this value. Three of the N_T measurements are faulty.

5.2. Detection of Faulty Measurements

We validate the algorithm, first, using the entire feature kinetic curve consisting of N_R values per measurement for the whole set of N_T measurements. We compare two different nominal P_{FA} w.r.t the actually obtained false alarm and correct detection rates. The results are presented in Table 1.

Observe that we are always able to correctly detect faulty frames. For a nominal $P_{\text{FA}} = 10^{-3}$ false alarms occur for very low glucose value measurements. They are confused with faulty measurements as the texture is quite similar. We see that the feature (M) shows the best performance w.r.t to achieving the nominal P_{FA}. This is due to the fact that under $\mathcal{H}_0$ the approximation of the residuals $r^i(n)$ by $f_{R^i} \sim \mathcal{N}(0, (\sigma_1^i)^2)$ is most accurate. For the features (E) and (C) more suitable models need to be found.

P_{FA}	Feature	False alarms rate	Correct detection rate
10^{-3}	E	1.8%	100%
	M	0.63%	100%
	C	2.5%	100%
10^{-4}	E	0%	100%
	M	0%	100%
	C	0%	100%

Table 1: Percentage of false alarms and correct detections for the features (E), (M) and (C) for the nominal false alarm rates $P_{\text{FA}} = 10^{-3}$ and $P_{\text{FA}} = 10^{-4}$.

5.3. Detection Time vs. Convergence Time

As longs as the reaction has not yet taken place ($t < t_0$) the features behave similarly for all measurements irrespective of them being faulty or not. Once the moistening period starts, the feature values quickly reach their final convergence value. Typically, t_0 lies around $1s$ depending on the current glucose measurement. In this experiment, we compute the test statistic sequentially for each new incoming frame

$$T_j(r^i(n)) = \frac{1}{N_j} \sum_{n=1}^{N_j} r^i(n) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda^i,$$

where N_j is the current frame. λ^i is calculated using a nominal $P_{\text{FA}} = 10^{-4}$. We detect the frame at which the measurement is detected as faulty. We perform this simulation for the three faulty measurements contained in our data set, which are denoted by K_1, K_2, K_3.

	E	M	C	t_C
K_1	$1.067s$	$1.067s$	$0.76s$	$3.33s$
K_2	$0.86s$	$0.967s$	$0.76s$	$3.0s$
K_3	$1.4s$	$1.3s$	$0.70s$	$5.33s$

Table 2: Detection times for the three faulty measurements in the real data set.

The correct measurements in the data set are all detected as being correct, as shown in the previous section. The results are listed in Table 2, as well as the convergence times t_C of the intensity values to serve as a comparison. Clearly, using our method we are able to detect the faulty frames long before the convergence times t_C. This means that the user does not have to wait until the device has calculated the estimate of the underlying glucose value to be told that the measurement is false but will get an error message after around $1s$ and can then repeat the measurement.

We can also observe that feature (C) converges quite quickly, giving the user a decision after $0.7 - 0.76s$.

The key point in this experiment is the convergence time of the features. For the faulty measurements we have proven

that they converge after the values given in Table 2. To circumvent the small sample size of faulty measurements and prove that this result will hold for larger sample sizes, we can detect the convergence times of the features for correct measurements. This was done and resulted in feature convergence times between $0.8s - 1.5s$. Hence, we are able to state that for any kind of measurement we will have a steady feature value closely after t_0 and will be able to detect whether the measurement is faulty or not.

6. CONCLUSIONS AND FUTURE WORK

We have developed a framework to detect faulty glucose measurements in hand-held devices that uses co-occurrence based texture analysis. We have identified the textural features energy, maximum probability and correlation to be suitable for detecting faults. Among the three, the feature maximum probability shows the best performance w.r.t to the false alarm rate. We have shown that the detection is achieved directly after the chemical reaction starts. The feature correlation converges the fastest to a steady value. For future work, we plan to further test our method with larger data sets containing more faulty measurements.

REFERENCES

[1] J.M. Asfour, H.P. Haar, G. Kocherscheidt, and W. Petrich, "Analyse Optischer Daten Mit Hilfe Von Histogrammen.," Dec. 07 2011, Europäische Patentanmeldung, EP1843148A1.

[2] N. Demitri and A.M. Zoubir, "Mean-shift Based Algorithm for the Measurement of Blood Glucose in Hand-Held Devices," in *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, September 2013.

[3] N. Demitri and A.M. Zoubir, "A Robust Kernel Density Estimator Based Mean-Shift Algorithm," in *IEEE 39th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2014.

[4] M. Tuceryan and A.K. Jain, *Handbook of Pattern Recognition and Computer Vision*, Singapore: World Scientific, 1993.

[5] R.M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactons on Systems, Man, And Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, November 1973.

[6] C. Chevrefils, F. Cheriet, C.E. Aubin, and G. Grimard, "Texture Analysis for Automatic Segmentation of Intervertebral Disks of Scoliotic Spines From MR Images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 608–620, April 2009.

[7] L.K. Soh and C. Tsatsoulis, "Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices," *IEEE Transactions on Geoscience And Remote Sensing*, vol. 37, no. 2,, March 1999.

[8] E. Baumann, W. Obermeier, and K. Werner, "Light source pulsed with irregular pulse sequence in analog photometric signal evaluation for a test carrier analysis system," Oct. 31 1995, US Patent 5,463,467.

[9] D.A. Clausi and H. Deng, "Design-based texture feature fusion using Gabor filters and co-occurrence probabilities," *IEEE Transactions on Image Processing*, vol. 14, no. 7, pp. 925–936, July 2005.

[10] S. Aksoy and R.M. Haralick, "Feature normalization and likelihood-based similarity measures for image retrieval," *Pattern Recognition Letters*, vol. 22, no. 5, pp. 563–582, April 2001.