Nucleus fingerprinting for the unique identification of Feulgen-stained nuclei

David Friedrich and Matthias Brozio and André Bell and Stefan Biesterfeld and Alfred Böcking and Til Aach

Institute of Imaging and Computer Vision
RWTH Aachen University, 52056 Aachen, Germany
tel: +49 241 80 27860, fax: +49 241 80 22200
web: www.lfb.rwth-aachen.de

in: SPIE Medical Imaging. See also BibTeX entry below.

BibTeX:

@inproceedings{FRI12b,
  author = {David Friedrich and Matthias Brozio and André Bell and Stefan Biesterfeld and Alfred Böcking and Til Aach},
  title = {Nucleus fingerprinting for the unique identification of Feulgen-stained nuclei},
  year = {2012},
  pages = {831517--831517},
  organization = {International Society for Optics and Photonics},
  journal = {SPIE Medical Imaging}
}

© copyright by the author(s)
Nucleus Fingerprinting for unique Identification of Feulgen-Stained Nuclei

David Friedrich\textsuperscript{a}, Matthias Brozio\textsuperscript{a}, André Bell\textsuperscript{a}, Stefan Biesterfeld\textsuperscript{b}, Alfred Böcking\textsuperscript{b} and Til Aach\textsuperscript{a}

\textsuperscript{a}Institute of Imaging and Computer Vision, RWTH Aachen University, 52056 Aachen, Germany
\textsuperscript{b}Department of Pathology, Division Cytopathology, Heinrich-Heine-University, 40204 Düsseldorf, Germany

ABSTRACT

DNA Image Cytometry is a method for non-invasive cancer diagnosis which measures the DNA content of Feulgen-stained nuclei. DNA content is measured using a microscope system equipped with a digital camera as a densitometer and estimating the DNA content from the absorption of light when passing through the nuclei. However, a DNA Image Cytometry measurement is only valid if each nucleus is only measured once.

To assist the user in preventing multiple measurements of the same nucleus, we have developed a unique digital identifier for the characterization of Feulgen-stained nuclei, the so called Nucleus Fingerprint. Only nuclei with a new fingerprint can be added to the measurement. This fingerprint is based on basic nucleus features, the contour of the nucleus and the spatial relationship to nuclei in the vicinity. Based on this characterization, a classifier for testing two nuclei for identity is presented.

In a pairwise comparison of \(\approx 40000\) pairs of mutually different nuclei, 99.5\% were classified as different. In another 450 tests, the fingerprints of the same nucleus recorded a second time were in all cases judged identical. We therefore conclude that our Nucleus Fingerprint approach robustly prevents the repeated measurement of nuclei in DNA Image Cytometry.

Keywords: DNA Image Cytometry, Nucleus Fingerprint, photo microscopy, cancer diagnosis

1. INTRODUCTION

DNA Image Cytometry (DNA-ICM) is a non-invasive method for cancer diagnosis and grading which measures the DNA content of nuclei. To this end, cell material is collected with a brush or a fine needle, fixated on a microscope slide and stained according to Feulgen. For this specific staining technique, the uptake of stain in a nucleus is proportional to its DNA content. A photomicroscopic system is used as a densitometer for measuring the amount of DNA in individual cells. The microscope slide is scanned until sufficient benign nuclei are collected for calibrating the system and enough suspicious nuclei are collected for diagnosis.\textsuperscript{1} Suspicious nuclei with exceedingly high DNA amount or stemlines which deviate significantly from a benign DNA distribution allow for a diagnosis and grading of cancer.\textsuperscript{2} DNA-ICM can be applied to various modalities, such as cell material gained by fine needle biopsies of the prostate or of serous effusions, or on brush smears from squamous epithelium, for instance of the oral or cervical cavity. As cancer is diagnosed on the cellular level, it is possible to detect cancer up to two-and-a-half-years earlier than with conventional histological methods.\textsuperscript{3} DNA Image Cytometry has recently been recommended to prevent overtreatment of prostate cancer.\textsuperscript{4,5}

For a valid DNA-ICM measurement, the consensus report of the European Society for Analytical Cellular Pathology prescribes that at least 300 morphologically suspicious nuclei and 30 reference nuclei for calibration have to be collected.\textsuperscript{1} If the user fails to collect sufficient cell material in the first scan, the slide has to be scanned again until the specifications of a valid measurement are met. However, repeated measurements of the same nucleus have to be avoided in any case, as it might bias the diagnosis or yield an invalid calibration of the

\textsuperscript{Further author information, send correspondence to David Friedrich, E-mail: david.friedrich@lfb.rwth-aachen.de, Telephone: +49 241 8027803.}
Fig. 1. Two different prostate nuclei, with the contour of the segmentation algorithm shown in blue. These nuclei have a similar contour and similar feature values.

System. Deciding whether a cell has been added to the measurement is prone to errors for a human observer, considering that the user performs many measurements on the same day and sees up to 75000 Feulgen-stained nuclei. Therefore, methods for ensuring the uniqueness of nuclei used in the measurement have to be developed. The decision could be facilitated if position information about the nuclei were available, however the required positioning systems are expensive and often not available. Alternatively, scene information could be stored and used for correlation based registration approaches, but this would require that the whole image scene for each nucleus collected is stored and that each new scene has to be compared with all scenes before. To overcome these problems, we compute a digital unique identifier for Feulgen-stained nuclei (Nucleus Fingerprint) which requires only little additional resources and for which the scene information is needed only during the time of the computation of the Nucleus Fingerprint. This Nucleus Fingerprint is based on basic cell features, its contour and the spatial relationship to nuclei in its vicinity. Combined with methods for comparing these identifiers, the Nucleus Fingerprint can be used as follows to prevent the multiple measurement of the same nucleus. When a nucleus candidate is to be added to the measurement, the Nucleus Fingerprint is computed. The fingerprint is compared to the fingerprints of all other nuclei in the measurement. Only when the Nucleus Fingerprint of the new nucleus is different from all fingerprints of already collected nuclei it is added to the measurement.

2. MATERIAL

A fine needle biopsy was used to extract prostate cell material from two patients. A Motic BA 410 microscope with 40x objective (NA=0.65) is used for imaging. The system is equipped with a Motic 285A camera (1360x1024 pixels, pixelsize $\Delta x = 0.18\mu m$). Images were shading corrected and nuclei were segmented by appropriate thresholding in RGB space, followed by morphological operations. Our training set consists of 20 field of views (FOVs) which do not overlap and thus include different cells (number of cells: 341). Subsequently, 32 nuclei from these FOVs were selected, moved to five different position in the field of view and acquired again at the new position. For the test set, the same procedure for acquiring nucleus image data was applied to an additional prostate slide (20 FOVs containing 282 cells, 45 nuclei acquired at 5 different positions).

We evaluate the Nucleus Fingerprint’s ability to distinguish different nuclei by testing all possible pairs of nuclei which were taken from non-overlapping FOVs and thus are mutually different. For $n$ cells, $\binom{n}{2}$ pairs can be compared, so 57970 mutually different nuclei can be compared in the training set and 39621 in the test set. On the same time, the Nucleus Fingerprint of the same nucleus acquired again at a different position of the FOV should be similar: This property is evaluated by comparing the fingerprints from the same nucleus acquired at different positions. As 32 and 45 nuclei were acquired five times at different positions for training and test set respectively, $32 \cdot \binom{5}{2} = 320$ and $45 \cdot \binom{5}{2} = 450$ comparisons of identical nuclei can be made.

3. NUCLEUS FINGERPRINT

To achieve a unique characterization yet using as little resources as possible, we reduce the feature space of a nucleus image by extracting basic nucleus features from the segmented nuclei as well as the contour of the segmentation mask. Features which are robust for the same nucleus, but have a high discriminative power in the pairwise comparison of different nuclei have to be extracted. Thereby, attributes of features which are usually
undesired in cell classification tasks, like variance with respect to the staining intensity or the rotation of the nucleus, can now provide valuable information. Still, Fig. 1 shows two different cells where the contour and basic cell features are similar. Consequently, the Nucleus Fingerprint cannot rely on the nucleus only, but the vicinity as a whole should be considered as well.

In the end, the comparison of two nuclei for identity is a binary decision and can be regarded as a two class classification problem (identical vs. different). In the following, we show how attributes extracted from basic nucleus features, the contour, and the vicinity of two nuclei can be used to calculate similarity measures \( F, C \) and \( V \) within the range \([0, 1]\). A weighted sum of these similarity measures is computed and the subsequent classification is performed by thresholding.

### 3.1 Basic nucleus features

Features are suitable for the Nucleus Fingerprint if they are, on the one hand, robust against noise when examining the same nucleus, and on the other hand, also discriminative for nuclei that are mutually different. We have computed a set of features from Rodenacker et al.\(^{10}\) and Holmquist et al.\(^{11}\) for all nuclei of the training set. To assess robustness against noise, we computed the variance of the feature values by acquiring their images multiple times. Discriminance was quantified by computing the coefficient of variation of feature values when computed for mutually different nuclei. Nine features showing a good performance in both aspects, that is low variance when comparing identical nuclei and high coefficient of variation when comparing mutually different nuclei, were selected: Area, perimeter, mean radius of the nucleus, variance of the nucleus radius, eccentricity, inertia, orientation, mean luminance and variance of the luminance values.

The numerical values of these features, however, span different ranges. To allow an uniform weighting of feature values, we therefore normalize the difference of two features. If two nuclei \( n_1 \) and \( n_2 \) are to be compared for identity, their feature values \( f_{k,n_1} \) and \( f_{k,n_2} \) of the feature \( k \in \{1, \ldots, 9\} \) can be normalized via

\[
F_{k,\text{norm}}(n_1, n_2) = \frac{|f_{k,n_1} - f_{k,n_2}|}{\frac{1}{2}|f_{k,n_1} + f_{k,n_2}|},
\]

which maintains symmetry of the criterion. The only exception among the features used is the orientation of the nucleus, which is defined as the deflection of the main axis to the x direction and yields values in the range \([0^\circ, 180^\circ]\). For this feature, its circular nature has to be considered. For instance, two nuclei which have an orientation of \(0.5^\circ\) and \(179.5^\circ\) almost point in the same direction. To account for this special case, this feature is compared with the formula

\[
F_{\text{orient, norm}}(n_1, n_2) = \min \left( |f_{\text{orient},n_1} - f_{\text{orient},n_2}|, 180 - |f_{\text{orient},n_1} - f_{\text{orient},n_2}| \right) \frac{90^\circ}{90^\circ}.
\]

The fraction of features which differ by less than 0.1 in the parameter \( F_{k,\text{norm}} \) is finally used as \( F \),

\[
F(n_1, n_2) = \frac{1}{9} \sum_{k=1}^{9} \mathbb{I}(F_{k,\text{norm}}(n_1, n_2) < 0.1),
\]

with \( \mathbb{I} = 1 \) if the inequality is true and \( \mathbb{I} = 0 \) otherwise.

### 3.2 Correlation of nuclei contours

A correlation-based approach is used for assessing similarity of the contours. The contour of a segmented nucleus can be represented as an \( N \)-tuple of Cartesian coordinates, when the boundary pixels of the nuclei are traversed in clockwise direction:

\[
( (x_l, y_l) \in \mathbb{R}^2 \mid l \in \{1, \ldots, N\}, x_l \in \{1, \ldots, 1024\}, y_l \in \{1, \ldots, 1360\} )
\]

Here, \( N \) is the length of the contour and the largest values for \( x_l \) and \( y_l \) originate from the camera resolution. Likewise, a complex representation

\[
\phi = (c_l \in \mathbb{C} \mid l \in \{1, \ldots, N\})
\]
of (4) is given by transferring the coordinates \((x_1, y_1)\) to their counterparts \(c_1\) in the complex plane. To compare the contour of nuclei which have been acquired at different positions of the FOV, we shift the contours such that the centroid of the nucleus is at the origin of the coordinate system. The coefficient of correlation of two contours \(\phi_1\) and \(\phi_2\) is then given by

\[
\text{corr}(\phi_1, \phi_2) = \frac{\langle \phi_1, \phi_2 \rangle}{||\phi_1||_2 \cdot ||\phi_2||_2} \in \mathbb{C}.
\] (6)

However, the contours from nuclei do not need to exhibit the same length, either because they have different size, or in the case of the same nucleus recorded again due to missegmentations or measurement noise. For the case that the contours have different length, the longer contour is cropped to the same length as the shorter one. Moreover, if the length of the contours differ too much (\(\mathcal{F}_{\text{perimeter, norm}} > 0.2\)), we regard this as a knock-out criterion against similarity and the contour criterion is set to zero. Otherwise, the complex correlation \(\text{corr}(\phi_1, \phi_2)\) is allocated as follows: Using the relationship between the scalar product \(\langle \cdot, \cdot \rangle\), the induced norm \(||\cdot||_2\) and the Cauchy Schwarz inequality,\(^{12}\) it can be shown that two identical contours yield a value of 1 if and only if \(\phi_1 = \alpha \phi_2\), with \(\alpha \in \mathbb{R}\) and \(\alpha > 0\). Therefore, the distance of \(\text{corr}(\phi_1, \phi_2)\) to 1 in the complex plane can be used as a criterion for similarity. This distance value is finally mapped to the range \([0, 1]\) and yielding 1 when the contours are identical with the formula

\[
C(\phi_1, \phi_2) = \max(0, 1 - |1 - \text{corr}(\phi_1, \phi_2)|).
\] (7)

3.3 Vicinity

As the stained nuclei have been fixated on the slide, the fact that the relative spatial relationship to nuclei in the vicinity remains constant is exploited by the following vicinity criterion: The position of a nucleus is robustly

Figure 2. Two FOVs where a nucleus has been acquired at different positions are superimposed. The joint vicinity in both FOVs with respect to this nucleus is marked by the dashed frame. The relative spatial relationship to nuclei in the vicinity is described by arrows.
described by its centroid, so the spatial relationships between the centroid \( n_j \) of a nucleus and the centroids of all nuclei \( n_{j,k} \) in the FOV describe its vicinity.

For comparing the vicinity of two nuclei \( n_1 \) and \( n_2 \), we first determine the number of neighbors which are within the joint viewing range of both nuclei. As these two nuclei might be taken at different positions in the field of view, only parts of the FOV might be visible for both nuclei (see Fig. 2). The viewing range of a nucleus is described by the distances from its centroid to the image borders in the left horizontal, right horizontal, upper vertical and lower vertical direction. The joint viewing range of two nuclei is given by the minimum visibility range distances of the two nuclei in each of these four directions. Then we determine the number of nuclei \( V_1 \) within this joint viewing range which have the same spatial relationship with respect to \( n_1 \) and \( n_2 \). In this context, a nucleus is at the same position if the centroids of the nuclei do not deviate by more than 0.5\( \mu \)m. The vicinity criterion \( V \) is then computed as the number of nuclei at the same position divided by the number of nuclei \( V_2 \) which are present in the joint viewing range of both FOVs,

\[
V = \frac{V_1}{V_2}.
\]

If no nuclei are within the joint viewing range of both nuclei, \( V_2 \) would be 0 and \( V \) cannot be computed by eq. (8). Consequently, for that case we define the criterion \( V \) to be invalid. Note that the case of an invalid vicinity criterion is different from a vicinity criterion with a return value of 0. In the latter case, nuclei are present in the joint viewing range, but they are not at the same positions.

Information which is necessary to generate the Nucleus Fingerprint and compare two vicinities are the visibility ranges and the spatial relationships to nuclei in the vicinity. These values are computed during the generation of the Nucleus Fingerprint and extracted from the image information of the current field of view, for comparing two Nucleus Fingerprints in a later stage of the measurement the image itself is not required. Deciding if two nuclei are identical can be achieved by first combining the similarity measures \( F \), \( C \) and \( V \) in a weighted sum, and apply a threshold \( T \) to test for identity:

\[
\omega_f F + \omega_c C + \omega_v V \leq_{\text{different}} T.
\]
Table 1. Classification table for the test set.

<table>
<thead>
<tr>
<th>class</th>
<th>truth</th>
<th>different</th>
<th>identical</th>
</tr>
</thead>
<tbody>
<tr>
<td>different</td>
<td>39422</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>identical</td>
<td>199</td>
<td>450</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>39621</td>
<td>450</td>
<td></td>
</tr>
<tr>
<td>error (%)</td>
<td>0.5</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

### 3.4 Decision for identity
On the training set, the vicinity feature proved to be a very powerful discriminator (see Sec. 4), being more powerful the more nuclei are visible. Therefore we increase the weight $\omega_v$ depending on the number of nuclei $V_2$ which are found in the joint viewing range via

$$\omega_v(V_2) = \frac{V_2}{V_2 + 12}.$$  \hspace{1cm} (10)

The weights $\omega_f$ and $\omega_c$ for the features and contour performed similar and thus their weights are set uniformly and such that the sum of weights is 1 using the formula

$$\omega_f(V_2) = \omega_c(V_2) = 0.5 \cdot (1 - \omega_v(V_2)).$$  \hspace{1cm} (11)

In the case that the vicinity criterion is invalid, $\omega_v(V_2)$ is set to zero and the weights for the feature and contour criterion to

$$\omega_f = \omega_c = 0.5.$$  \hspace{1cm} (12)

### 4. RESULTS
In a first step, the similarity measures from basic features, the contour and the vicinity were analyzed individually on the training set. On the one hand, the similarity criteria should yield values close to zero when computed for mutually different nuclei. Average values were $F = 0.15$, $C = 0.25$. Different from the criteria $C$ and $F$, which can be computed in all cases, the vicinity criterion is only valid if the joint viewing range contains at least one nucleus (Sec. 3.3). This was possible in 98% of the cases, then the average value was $V = 7 \cdot 10^{-4}$. On the other hand, these three criteria should robustly give values close to one when computed for the same nuclei recorded at different positions, here the average values were $F = 0.95$, $C = 0.94$ and $V = 0.97$ (available in 93% of the cases).

Secondly, the overall performance of the Nucleus Fingerprint was examined: Fig. 3 shows the normalized histograms of the Nucleus Fingerprint values computed as in (9) for identical and different nuclei from the training set. This plot reveals that both classes can be well separated by thresholding. Based on this analysis, a threshold of $T = 0.7$ was chosen for evaluating the test set.

Finally, the classification results on the test set are shown in Table 1. From the mutually different nuclei, 99.5% were correctly classified. For the identical nuclei, all nuclei were correctly classified as such. An analysis of misclassifications of mutually different nuclei revealed that for 86.9% of these misclassifications, the powerful vicinity criterion is not available, whereas it is only missing for 2% of all pairs of mutually different nuclei.

### 5. CONCLUSIONS
We have presented a framework for comparing two Feulgen-stained nuclei based on unique and robust Nucleus Fingerprints. This framework can be used in the scope of a DNA Image Cytometry measurement to prevent multiple measurements of the same nucleus. The total misclassification error was 0.5%, and the type of misclassification would only prevent that a small fraction of nuclei can not be added to the measurement, while it would still comprise only unique nuclei. The Nucleus Fingerprint requires very little additional memory for each
nucleus, as only nine feature values, the contour, and the distances to nuclei in the vicinity have to be stored. Likewise, the comparison of two cells has been reduced to the comparison of few but selected parameters.

The misclassified pairs of nuclei contain a significantly high fraction of pairs where the vicinity criterion is not available. We assume that the misclassifications can be further reduced with minimal additional user interaction: If the vicinity criterion is not available but the other similarity measures take high values when compared to a specific nucleus \( n \), the user can be asked to acquire the nucleus again at a position close to where \( n \) was acquired. Additionally, both nuclei can be shown to the user for a final decision. Future research will include adapting the Nucleus Fingerprint to other materials, like cells originating from oral or cervical smears or extracted from pleural effusions.

6. ACKNOWLEDGEMENTS

The financial support as well as the provision of technical equipment from Motic China Group Co., Ltd., Peoples Republic of China, is gratefully acknowledged.

REFERENCES


