# Automated quantification of histone relocation in cell nuclei

T. Rieß[1], C. Dietz[2,1], M. Horn[3,1], O. Deussen[4,1], M. Leist[5], T. Waldmann[5] & D. Merhof[6,1]

[1] *Interdisciplinary Center for Interactive Data Analysis, Modelling and Visual Exploration (INCIDE), University of Konstanz, Germany*

[2] *Center for Applied Photonics (CAP), University of Konstanz, Germany*

[3] *Bioinformatics and Information Mining, University of Konstanz, Germany*

[4] *Computer Graphics and Media Design, University of Konstanz, Germany*

[5] *In-vitro Toxicology and Biomedicine, University of Konstanz, Germany*

[6] *Visual Computing, University of Konstanz, Germany*

ABSTRACT: Microscopy images of progenitor cells prepared with a specific epigenetic mark show its massive relocation during differentiation. We propose an image processing pipeline for the automated quantification of this relocation. A novel set of features which are extracted from polar representations of the microscopy images are used to construct a strong classifier via an adaptive boosting algorithm. In addition to the classification a measurement of the relocation in each nucleus is performed.

Keywords: biomedical image processing, feature extraction, image classification

## 1 INTRODUCTION

The last step during differentiation of mammalian tissues is the maturation phase when cell type specific gene expression are switched on to differentiate the progenitor cells into a specific cell type (Zimmer, Kuegler, Baudis, Genewsky, Tanavde, Koh, Tan, Waldmann, Kadereit, and Leist 2011). This switch is not only regulated by linear DNA sequence information, but also by the structural organization of DNA, called chromatin (Khorasanizadeh 2004). In this work we propose an image processing pipeline to quantify the chromatin structural changes during the maturation phase of the differentiation.

The microscopy images under consideration show immunostained progenitor cells at different days of differentiation with a well described repressive histone epigenetic mark (Kouzarides 2007). This mark is massively relocated during differentiation, from a relative homogenous and regular pattern toward a clear concentration at the nuclear periphery. To show that this relocation is not only occurring in a subset of cells, we quantify the amount of cells with ring structure at the nuclear periphery during differentiation for a large number of microscopy images, see Figure 1 for two exemplary images.

Our image processing pipeline automatically classifies the microscopy images using a novel set of rotation-invariant features and an adaptive boosting algorithm. Moreover, for comparison the histone relocation is measured using an additional fluorescent marker. The methods are presented in detail in Section 2, the results of the automated classification and the measurement are summarized in Section 3, and finally in Section 4 conclusions are drawn and future research directions are given.
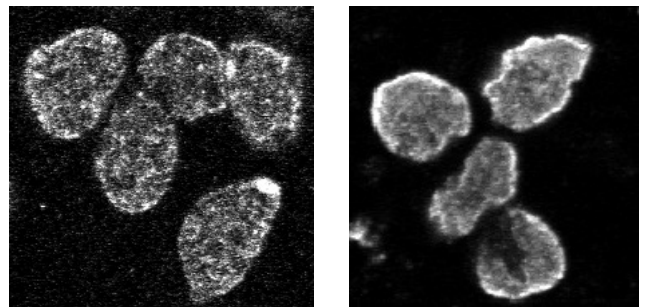


Figure 1: Cell nuclei after two days (left) and after seven days (right) of differentiation. The cells show a clear ring structure after seven days.

## 2 METHODS

In this section we present the technical details of the proposed image processing pipeline for the automated classification of microscopy images. In particular, we explain which features are extracted and used for the adaptive boosting algorithm.

### 2.1 *Feature extraction*

The features are extracted from polar representations (Kvarnström, Logg, Diez, Bodvard, and Käll 2008) created from the original microscopy image. These so-called polar images are determined by choosing a center point, a radius and a number of angles. Then rays originating at the center point and having the length of the chosen radius are collected and stacked for the chosen number of angles, see Figure 2 for an example. The polar images we use here are centered at seeding points that are distributed on a rectangular grid on the microscopy image. The distance between neighboured seeding points is chosen to be $r/5$, where $r$ denotes the average nucleus radius in our test images. We use $4r$ angles and a radius of $6r/5$ for the construction of the polar images, thus each polar image has the width $W = 6r/5$ and the height $H = 4r$. Using our specific set of features on polar images yields rotational invariance, but not necessarily scaling invariance.

The features explained below are not only extracted from the complete polar image, but also from subwindows of the polar image. These subwindows cover the whole range of angles, meaning that their height is $H$. Hence, the rotational invariance of the features is preserved. We choose subwindows of width $W/2$ centered at $x$-positions $W/4$, $W/2$ and $3W/4$ within the polar image to extract the features. Note that the total number of features is proportional to the number of subwindows. In the following detailed explanation of the extracted features, let $w$ and $h$ denote the width and the height of the considered subwindow.
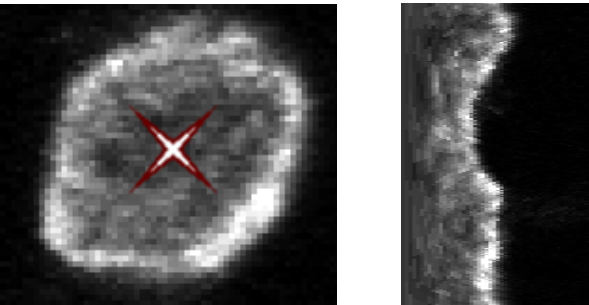


Figure 2: Polar image for a single cell nucleus. The microscopy image (left) is converted to a polar image (right). The center of the polar image is chosen close to the center of the nucleus.

### *Histogram features*

The histogram of the subwindow is computed and the following statistical values are used as features: mean, median, $20\%$-quantile, $80\%$-quantile, standard deviation and kurtosis. Histogram features describe global statistics of the pixel intensities.

### *Haar-like features*

The response of a Haar-like feature (Lienhart, Kuranov, and Pisarevsky 2003) measures the difference of pixel intensities in distinct rectangles. Recall that the considered subwindow of the polar image is $w$ pixels wide and $h$ pixels high. Let $(a, b, c, d)$ denote a rectangle that is defined by its upper left corner $(a, b)$ and its lower right corner $(c, d)$.

The first set of values are computed using two rectangles $(0, y, w/2 - 1, y)$ and $(w/2, y, w - 1, y)$ for $0 \leq y < h$. This means that the pixel intensity of the right half of each row $y$ in the subwindow is subtracted from the pixel intensity of the left half, see Figure 3 (left) for a sketch.

The second set of values are computed using a rectangle $(0, y, w - 1, y + h/2)$ for $0 \leq y < h/2$. The pixel intensity within this rectangle is compared to the pixel intensity of the remaining area of the subwindow, see Figure 3 (right) for a sketch.

For both sets of computed values, we again use mean, median, $20\%$-quantile, $80\%$-quantile, standard deviation and kurtosis as the actual features. Both of these feature sets are a measure for how likely the seeding point is inside a cell nucleus, and how close this point is to the center of the nucleus. In particular, the second Haar-like feature is a measure for the symmetry.
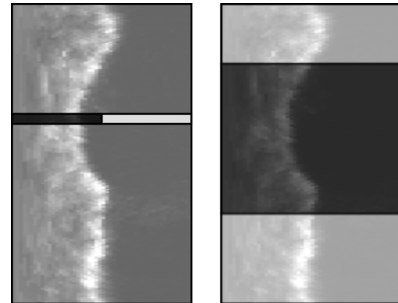


Figure 3: Sketch of the used rectangles for the Haar-like features. The left panel illustrates the per-row comparison of the left half and the right half, the right panel shows the comparison of a rectangle with full width and half height versus the remaining part.

*Shape features*

In order to extract the shape features, first a Gaussian filter with radius 3 is applied to the polar image. Then a Haar-like filter is applied to this image in the following manner: The value of the pixel at position $(x, y)$ is computed as the difference of the pixel intensity of rectangle $(x - 4, y, x - 1, y)$ and the pixel intensity of rectangle $(x, y, x + 3, y)$. This value is then divided by $\ln x$. Afterwards, a dynamic programming approach is used to detect a vertical contour of maximum responses in this filtered image. The advantage of this contour detection algorithm over a standard gradient filter is that it is much more robust to noise. Moreover, it is possible to take the direction of the transition into account. Figure 4 shows a Haar-like filtered polar image and the resulting contour.

Let $c(y)$ denote the $x$-coordinate of the contour in the polar image. The response of a Haar-like feature comparing two rectangles $(c(y) - 4, y, c(y) - 1, y)$ and $(c(y), y, c(y) + 3)$ is computed for each row $y$ in the original polar image. Similarly, the pixel intensity, its median, mean, sum, maximum and minimum within the rectangle $(c(y) - 4, y, c(y) + 3, y)$ is computed in the original polar image.

The actual extracted features are again the statistical values mean, median, 20%-quantile, 80%-quantile, standard deviation and kurtosis of the computed values. These features are measures for the shape and the intensity of the contour.
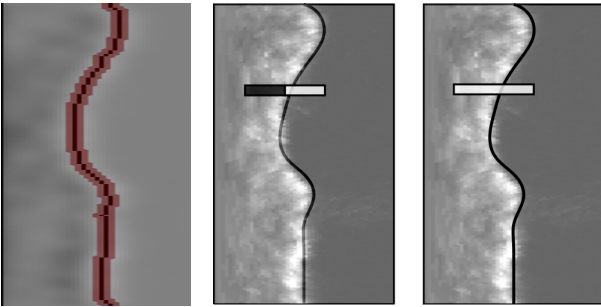


Figure 4: Haar-like filtered polar image (left), the computed contour is overlaid in red. The middle panel shows the original polar image overlaid with a sketch of the contour and the rectangles used for the Haar-like feature response in each row. The right panel shows a sketch of the rectangle used for the intensity measurements.

## 2.2 *Boosting*

We use an adaptive boosting algorithm Ada-Boost (Freund and Schapire 1999) to classify the polar images. The main idea of the AdaBoost algorithm is to use a linear combination of many weak classifiers in order to construct a strong classifier, where the weak classifiers are simple decision stumps using the extracted feature values, see (Smith, Carleton, and Lepetit 2009; He, Wang, Metaxas, Mathew, and White 2007; Viola and Jones 2004) for similar approaches. At the beginning of the boosting process, all weak classifiers are equally weighted. The boosting algorithm uses training data to adjust the weights of the weak classifiers according to the distinctiveness for the class membership. This weight adjustment process is repeated iteratively until either a maximum number of iterations is reached, or the global classification error of the training data is below a given threshold.

For our classification of polar images, we use this algorithm twice. In a first boosting step, we construct a strong classifier for the decision between the two classes "nucleus" and "background", which distinguish if the seeding point of the polar image is inside or outside a cell nucleus. Only for some of the polar images classified as "nucleus", a second boosting step is performed, which constructs a strong classifier for the distinction of the classes "ring" and "no ring". This classifier determines if the seeding point belongs to a nucleus with or without a visible ring structure which is due to the histone relocation. To further improve the performance in the first boosting step "nucleus" versus "background", we also use features extracted from a second channel of the microscopy images, which only shows the DAPI stained nuclei.

## 2.3 *Classification*

Using the strong classifiers constructed by the boosting phase explained above, unlabelled microscopy images can be classified. Prior to the classification, all features of all polar images are extracted. Similar to the two steps of the boosting algorithm, the classification consists of two stages: First, the classification of "background" versus "nucleus" is performed, which also uses the DAPI channel as mentioned above. The seeding points classified as "nucleus" are then pruned by keeping only those that have maximum response within a circle of radius $r$ (this means that only one seeding point per nucleus is kept). The polar images corresponding to those seeding points are then classified into the two classes "ring" and "no ring". Note that the whole process does not involve any segmentation, it is a pure classification of polar images created from the original microscopy image. In order to improve the classification performance, we use additional images: The images are acquired via confocal microscopy, thus there is a complete image stack available. Using a combination of the classification results of the three images in the middle of the stack yields our final classification result.

## 2.4 *Direct measurement of the histone relocation*

In addition to the classification explained above, the histone relocation per nucleus can be measured directly. For this measurement, an additional fluorescent marker (lamin) is applied to the experiment, which marks the periphery of the cell nuclei. This information is then extracted from a third channel in the microscopy images, see Figures 5, 6 and 7 for the three channels of an example image. The image processing pipeline for the measurement then consists of the following steps:
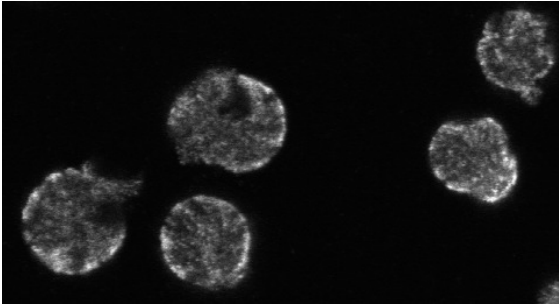


Figure 5: The histone channel of an example image (here: after 4 days of differentiation). The intensities are measured (per nucleus) in this channel.
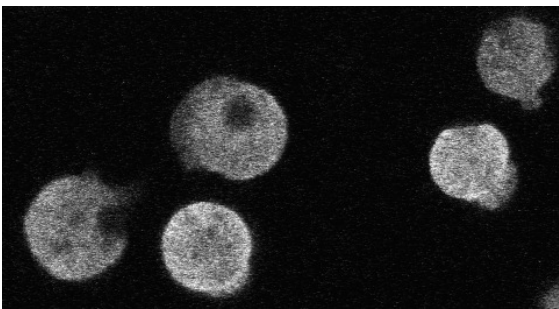


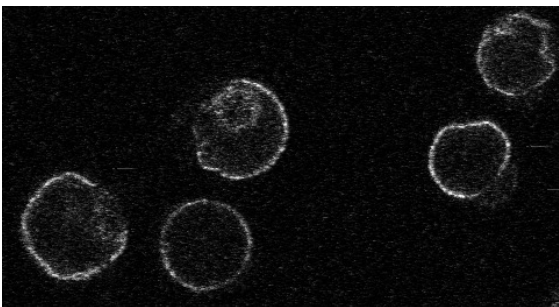Figure 6: The DAPI channel of an example image. This channel is used to segment the nuclei.



Figure 7: The lamin channel of an example image. This channel is used to segment the periphery of the nuclei.

### Nucleus segmentation

On the DAPI channel, single cell nuclei are segmented using a standard thresholding algorithms in conjunction with a connected components algorithm and basic morphological operations (erode, dilate). See Figure 6 for the DAPI channel of an example image.

### Periphery segmentation

On the lamin channel, again standard thresholding, connected components and morphological operations are used to segment the periphery of the cell nucleus. See Figure 7 for the lamin channel of an example image and Figure 8 and Figure 9 for the resulting segmentations of the periphery and the inner part.

### Intensity measurement

On the histone channel, the actual intensity on a ring along the periphery and in the inner part of each nucleus is measured. The ratio of the intensity of the inner part of each nucleus and the periphery is recorded. See Figure 5 for the histone channel of an example image.
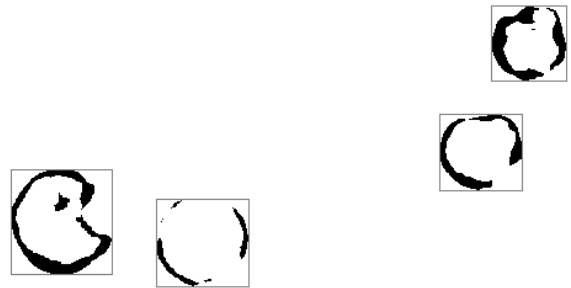


Figure 8: Segmentation of the nucleus periphery from the lamin channel of the example image, see Figure 7. The intensity along this periphery is measured in the histone channel, see Figure 5.
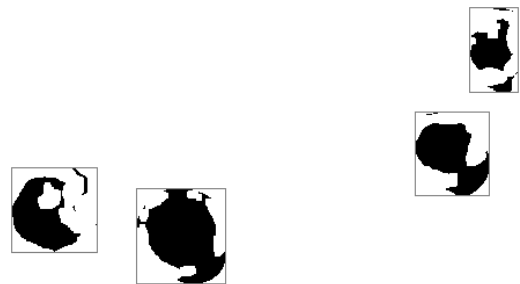


Figure 9: Segmentation of inner part from the lamin channel, see Figure 7. The intensity in the inner part is measured in the histone channel, see Figure 5.

## 2.5 *Implementation*

Both image processing pipelines, for classification of histone relocation and for the direct measurement, are implemented in the data processing framework KN-IME (Berthold, Cebron, Dill, Gabriel, Kötter, Meinl, Ohl, Sieb, Thiel, and Wiswedel 2008), which is a general tool to algorithmically process data in a very convenient way. A KNIME-workflow consists of interconnected nodes, each of which can either load data, process data or output data, while the data is passed from one node to another via interactively defined connections. Recent extensions provide all necessary image processing algorithms that can readily be applied to microscopy images as presented in this work.

## 3 RESULTS

We use a set of 46 images with a total of 1990 classified seeding points as ground truth (1150 "background", 840 "nucleus" — 250 "ring", 590 "no ring"). For the first boosting "nucleus" versus "background", we use 300 random seeding points for each class as training data, for the second boosting "ring" versus "no ring", we use 250 random seeding points for each class. The linear combination of weak classifiers is slightly biased towards "background" in order to avoid too many false positives in the first boosting step, and it is slightly biased towards "ring" in the second boosting step to improve the recognition rate for weak rings.

The global classification error in the boosting phase is shown in Figure 10. The boosting "nucleus" versus "background" only uses 17 features and reaches a global error of 0.01, while the boosting "ring" versus "no ring" uses 50 features and reaches a global classification error of 0.054. Table 1 shows which features are actually used in each boosting.
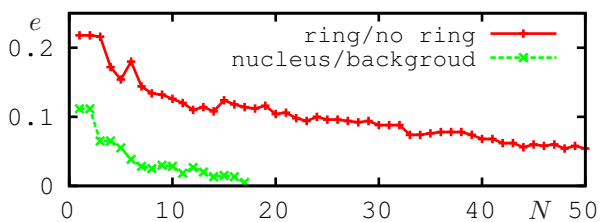


Figure 10: Number of used features $N$ versus global classification error $e$ in the boosting phase. The stop criterion is $e < 0.01$ or $N > 50$.

| Boosting | # hst | # Haar-l. | # cont. |
|---|---|---|---|
| Nucleus/background | 1 | 8 | 8 |
| Ring/no ring | 11 | 12 | 27 |

Table 1: Number of histogram (hst), Haar-like and contour features used in the two boosting steps.

The final classification result of 29 microscopy images of progenitor cells and at different days of differentiation are shown in Table 2, Figure 11 shows an exemplary image. The first classification step (nucleus detection) has a total detection rate of 97.0%. The classification results are as expected, the ring structure is detected in almost all cells after 7 days. Note that the poor classification result on day 4 is due to the fact that at this stage the histone relocation is not yet completed, which yields a very difficult classification.

| # | day 2 | day 4 | day 7 |
|---|---|---|---|
| no r. | 467 (95.7%) | 134 (67.3%) | 13 (8.0%) |
| ring | 21 (4.3%) | 65 (32.7%) | 149 (92.0%) |
| fp n.r. | 2 (0.4%) | 20 (14.9%) | 2 (15.4%) |
| fp r. | 12 (57.1%) | 19 (29.2%) | 1 (0.7%) |

Table 2: Results for images of progenitor cells. Shown are the number of nuclei classified as "ring" and "no ring" on the days of differentiation, and the number of false positives (fp) for both classes.
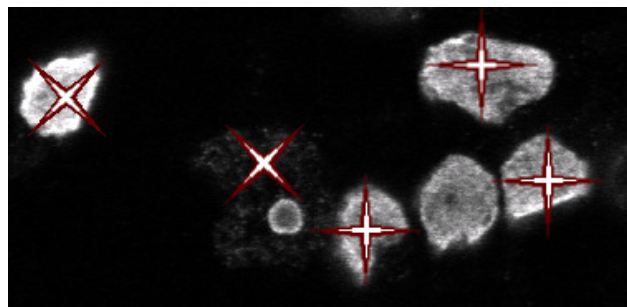


Figure 11: Example classification result. Nuclei classified as "ring" are marked by "X", nuclei classified as "no ring" are marked by "+". Note that the algorithm misses one nucleus and there is one false positive "ring" in this image.

For comparison, the direct measurement results are presented in Figure 12. Shown are the mean of the ratios of the intensities of the histone signal along the periphery and in the inner part of the nucleus along with the standard error bars. Note that the measurement are very sensitive to the image quality, and in the presented experiments the degradation of the lamin marker was an issue, hence the measurement results for day 7 only use a subset of the original image set in which the marker is still present (but this decreases the measurement accuracy).

## 4 CONCLUSIONS

We presented an image processing pipeline for an automated classification of cell nuclei in microscopy images. The classification is based on features extracted from polar images that are weighted using an adaptive boosting algorithm. In addition to the classification,
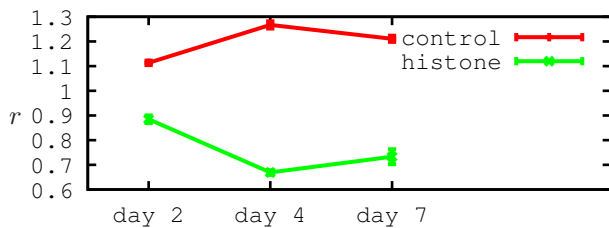
Figure 12: Direct measurement results using the ring segmentation and intensity measurement per cell. The day of the experiment is plotted versus the mean of the intensity ration $r$ (the intensity of the inner part of the nucleus divided by the intensity along the periphery per cell). Shown are the results for histone and for a control marker, which is known not to accumulate at the nucleus periphery. Note that for the histone measurement of day 7 only a small subset of the images were usable due to degrading of the marker.

we presented a second image processing pipeline for direct measurement of the histone relocation per cell nucleus. Both approaches have clear advantages and disadvantages: The classification works very well for clearly visible relocation or non-relocation, for example at the beginning or at the end of the experiment, and it is relatively robust to poor image quality. However, it does not perform optimal in the middle of the experiment, where the relocation is not yet distinct. On the other hand, the direct measurement works well in each stage of the experiment, but it is very dependent on the correct segmentation and on the image quality, in particular the degrading of the lamin marker decreases the percentage of correct measurements.

Future research directions involve a reduction of the number of labelled training images necessary for a robust classification, and the usage of extended Ada-Boost variants for multi-class classification, in particular to improve the classification results of images showing cells without a clearly visible ring structure. Moreover, an improvement of the direct measurement method using more advanced segmentation methods is desirable.

## 5 ACKNOWLEDGEMENTS

REFERENCES

Berthold, M., N. Cebron, F. Dill, T. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, and B. Wiswedel (2008). KNIME: The Konstanz Information Miner. In *Proc. Data Analysis, Machine Learning and Applications*, pp. 319–326.

Freund, Y. and R. Schapire (1999). A short introduction to boosting. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pp. 1401–1406. Morgan Kaufmann.

He, W., X. Wang, D. Metaxas, R. Mathew, and E. White (2007). Cell segmentation for division rate estimation in computerized video time-lapse microscopy. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Volume 6431.

Khorasanizadeh, S. (2004). The nucleosome: from genomic organization to genomic regulation. *Cell* (116), 259–272.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* (128), 693–705.

Kvarnström, M., K. Logg, A. Diez, K. Bodvard, and M. Käll (2008). Image analysis algorithms for cell contour recognition in budding yeast. *Optics Express 16*(17), 12943–1257.

Lienhart, R., E. Kuranov, and V. Pisarevsky (2003). Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In *DAGM 25th Pattern Recognition Symposium*, pp. 297–304.

Smith, K., A. Carleton, and V. Lepetit (2009). Fast Ray features for learning irregular shapes. In *IEEE 12th International Conference on Computer Vision*, pp. 397–404.

Viola, P. and M. Jones (2004). Robust real-time face detection. *International Journal of Computer Vision 57*, 137–154.

Zimmer, B., P. Kuegler, B. Baudis, A. Genewsky, V. Tanavde, W. Koh, B. Tan, T. Waldmann, S. Kadereit, and M. Leist (2011). Coordinated waves of gene expression during neuronal differentiation of embryonic stem cells as basis for novel approaches to developmental neurotoxicity testing. *Cell death and differentiation* (18), 383–395.