

Automated analysis of time-lapse fluorescence microscopy images: from live cell images to intracellular foci

Oleh Dzyubachyk¹, Jeroen Essers^{2,3}, Wiggert A. van Cappellen⁴, Céline Baldeyron², Akiko Inagaki⁴, Wiro J. Niessen^{1,5} and Erik Meijering^{1,*}

¹Biomedical Imaging Group Rotterdam, Departments of Medical Informatics and Radiology, ²Department of Cell Biology and Genetics, ³Departments of Radiation Oncology and Vascular Surgery, ⁴Department of Reproduction and Development, Optical Imaging Centre, Erasmus MC, Rotterdam, The Netherlands, ⁵Imaging Science and Technology, Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands

Associate Editor: Alex Bateman

ABSTRACT

Motivation: Complete, accurate and reproducible analysis of intracellular foci from fluorescence microscopy image sequences of live cells requires full automation of all processing steps involved: cell segmentation and tracking followed by foci segmentation and pattern analysis. Integrated systems for this purpose are lacking.

Results: Extending our previous work in cell segmentation and tracking, we developed a new system for performing fully automated analysis of fluorescent foci in single cells. The system was validated by applying it to two common tasks: intracellular foci counting (in DNA damage repair experiments) and cell-phase identification based on foci pattern analysis (in DNA replication experiments). Experimental results show that the system performs comparably to expert human observers. Thus, it may replace tedious manual analyses for the considered tasks, and enables high-content screening.

Availability and implementation: The described system was implemented in MATLAB (The MathWorks, Inc., USA) and compiled to run within the MATLAB environment. The routines together with four sample datasets are available at <http://celmia.bigr.nl/>. The software is planned for public release, free of charge for non-commercial use, after publication of this article.

Contact: meijering@imagescience.org

Received on June 2, 2010; revised on July 19, 2010; accepted on July 23, 2010

1 INTRODUCTION

The ability to perform analyses on individual cells presents evident advantages over the traditional averaging over the whole cell population (Gordon *et al.*, 2007). In many studies, such analyses are mainly performed manually, which is very tedious and often lacks accuracy, completeness and reproducibility. To improve this, automated methods are essential. In this article, we present a new system for performing intracellular analysis in time-lapse fluorescence microscopy image data of cell colonies. The system consists of two main modules: cell analysis and foci analysis. The first is more generic and can be applied to a large variety of biological data acquired for cell analysis. The second is naturally a

more application-dependent step and requires specialized methods depending on the structures of interest.

A number of cell segmentation and tracking algorithms have been presented in the recent literature (Al-Kofahi *et al.*, 2006; Dufour *et al.*, 2005; Dzyubachyk *et al.*, 2010 and the references therein; Ersoy *et al.*, 2009; Li *et al.*, 2008; Padfield *et al.*, 2009). However, very few can potentially satisfy the requirements imposed by live cell imaging and analysis at the individual cell level. Specifically, a candidate algorithm should have the capability to handle 3D time-lapse image datasets, it should provide full segmentation (detection only is insufficient) and tracking, be able to handle cell divisions, and show good performance even in the presence of significant noise and inhomogeneous intensity distributions (whether in the background or within cells). To this end, we use our robust level-set-based cell segmentation and tracking algorithm (Dzyubachyk *et al.*, 2010) as a starting point. Here, we present an extension of the algorithm that allows registration of each cell to a common coordinate system by applying motion correction after segmentation and tracking. This is necessary to study the true relative dynamics of intracellular processes.

As for the subsequent step of intracellular analysis, we focus here on fluorescent foci, which appear in many biological studies. Representing high concentrations of the corresponding fluorescently labeled protein, foci are usually the main indicator of an underlying biological process occurring at these locations (Gerlich and Ellenberg, 2003; Leonhardt *et al.*, 2000). Consequently, this makes foci analysis the main tool for studying protein-related processes by means of fluorescence microscopy. Examples of biological research based on foci analysis include fluorescent *in situ* hybridization (FISH) experiments (Gué *et al.*, 2005; Kozubek *et al.*, 1999; Netten *et al.*, 1996; Raimondo *et al.*, 2005), analysis of DNA replication and repair (Essers *et al.*, 2005a; Inagaki *et al.*, 2009; Meister *et al.*, 2007), and classification of cell-cycle phases (Ersoy *et al.*, 2009). In this article, we present a novel foci segmentation algorithm, and evaluate its robustness in segmenting foci of different size and intensity, as well as clustered foci.

In addition to presenting the algorithms used in different steps, we also validate the complete system by showing its ability to reproduce findings from two biological studies that were based on expert manual analyses. In the first experiment, we investigate the time course of nuclear foci formation and disappearance upon treatment with ionizing radiation of the 53BP1 DNA repair protein.

*To whom correspondence should be addressed.

In the second experiment, we employ the system for identifying cell phases in time-lapse images of proliferating cell nuclear antigen (PCNA)–green fluorescent protein (GFP)–stained cells. PCNA is a central protein in DNA replication and PCNA foci mark the sites of active DNA synthesis. Thereby, automated cell-phase identification is an important application that will facilitate further research of cell-cycle-related studies (Sigal *et al.*, 2006), in particular, cancer drug discovery (Wang *et al.*, 2008). To this end, as part of the second step of the system, we developed a simple yet effective algorithm for cell-phase detection based on observed typical PCNA foci patterns through the cell cycle (Leonhardt *et al.*, 2000). The results of the validation experiments clearly show the potential of the system for performing screening of high-content cell-based assays in applications involving the considered tasks.

The main contribution of this work is that we combine cell analysis and foci analysis algorithms into a single, fully automated system. In addition, we propose novel extensions and improvements to each of the components. We have recently shown (Dzyubachyk *et al.*, 2010) that our level-set-based cell segmentation and tracking algorithm is more accurate and robust than other state-of-the-art methods, especially in image sequences with greatly varying object intensity distributions. Using this algorithm as a basis, we have applied a shape-based motion correction algorithm, which to our knowledge has not been used before for this purpose. We also present a novel algorithm for foci segmentation, which is able to segment foci of different sizes, geometries and intensities, as well as clustered foci. Based on this, we propose a novel approach to cell-phase identification, which utilizes features from the segmented foci, rather than from the raw (noisy) images. Our system can be potentially used for any biological application requiring combined cell and foci analysis.

2 METHODS

The developed system processes images in a top–down fashion: (i) cell analysis and (ii) foci analysis. Here, we present the methods developed for performing these tasks.

2.1 Cell analysis

To prepare for analysis of intracellular structures, it is necessary to first determine the position and outline of each cell in the image data. Often it is also useful to transform the found cells to a common coordinate system to analyze intracellular changes free of global cell motion. This requires two processing steps: (i) cell segmentation and tracking; and (ii) cell motion correction.

2.1.1 Cell segmentation and tracking Segmentation and tracking of cells in image sequences is a difficult task. Especially, in live cell-imaging experiments it is hampered by low signal-to-noise ratio, cell clustering (unclear cell boundaries), inhomogeneous intensity distributions (in the background or within the cells) and intensity decay (due to photobleaching). In this system, we have adopted our recently developed level-set-based cell segmentation and tracking algorithm (Dzyubachyk *et al.*, 2010). The algorithm performs simultaneous segmentation and tracking by means of a model evolution approach, employing level sets as the underlying model. In the cited paper, we have shown that such approach guarantees a high quality of segmentation under strongly varying object intensities (whether spatially or temporally), the ability to handle data of any dimensionality (2D, 3D or even higher) without requiring fundamental changes to the algorithm, and natural handling of topological changes, which is a prerequisite when dealing

with dividing cells. A detailed description of this algorithm can be found in the cited paper.

2.1.2 Cell motion correction Motion correction methods can be roughly divided into two groups: feature and shape based (or area based) (Zitova and Flusser, 2003). The former use information about image features (usually related to image intensity), whereas the latter use shape information only. The choice for one type or the other is dependent on the underlying biological application. Both types of methods have been applied successfully for motion correction of segmented cells (Kim *et al.*, 2007; Mattes *et al.*, 2006; Matula *et al.*, 2006; Yang *et al.*, 2008). However, none of these methods can be applied directly to our problem. First, since our ultimate aim is to perform analysis of intracellular structures, only shape-based registration can be used. Second, the method should be able to separate global cell motion from local deformations, which is an ill-posed problem (Yezzi and Soatto, 2003).

To solve this problem, we have adapted the approach of Paragios *et al.* (2003), where a shape is described by a signed distance function. This perfectly fits our needs, as the output of the cell segmentation and tracking step are level-set functions using the same representation. Shape registration is then performed via energy minimization, using an energy functional that contains terms representing both global motion and local deformation. Since normally cells do not change shape dramatically between two consecutive time steps, we register each image to its predecessor. The only exception is cell mitosis, during which a cell undergoes considerable (and quite typical) shape change. To deal with such cases, we consider the newly born daughter cells as new objects, and initiate a new registration sequence for each of them. Thus, registration is performed on the full lifespan of a cell: from the moment after division (or from the first frame in the sequence) until the moment when the cell divides (or until the last frame).

Another issue arises from the typical sparseness of microscopy data along the z -axis in 3D. As pointed out by Matula *et al.* (2006), the rotation of cells in a typical assay is virtually limited to rotation around the z -axis only, and since vertical displacement is practically absent too, the registration task essentially becomes a 2D problem. Therefore, we perform registration on the maximum intensity projection of the 3D cell region, and apply this transformation to each slice of the 3D image.

In our algorithm, the deformation of a 2D cell region is described by rotation angle θ , shift $T = (T_x, T_y)$, scaling factor s , and local deformation field (U, V) . Shape registration is then achieved by minimization of the following energy functional:

$$\begin{aligned} E(s, \theta, T, (U, V)) = & \alpha \iint_{\Omega} N_{\delta_1} (s\Phi_D - \Phi_S(A))^2 d\Omega \\ & + (1 - \alpha)\beta \iint_{\Omega} N_{\delta_2} (s\Phi_D - \Phi_S(A - (U, V)))^2 d\Omega \\ & + (1 - \alpha)(1 - \beta) \iint_{\Omega} N_{\delta_2} (U_x^2 + U_y^2 + V_x^2 + V_y^2) d\Omega, \end{aligned} \quad (1)$$

where Φ_D and Φ_S are the signed distance functions corresponding to the source and the target shapes, Ω is the image region, $N_{\delta_1} = N_{\delta_1}(\Phi_D, \Phi_S)$ and $N_{\delta_2} = N_{\delta_2}(\Phi_D, \Phi_S)$ are narrow bands around the shape contours and $\alpha, \beta \in [0, 1]$ are balancing weights,

$$A(x, y) = \begin{pmatrix} A_x \\ A_y \end{pmatrix} = s \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \end{pmatrix} \quad (2)$$

is the image transformation, and (x, y) are the Cartesian coordinates on Ω . Here, the non-rigid deformation field (U, V) serves as a complement to the transformation A to ensure better fitting and convergence. For generating the warped image, either only rotation and shift, or all the registration parameters are used, depending on the application. This way, the global motion of the object (in the first case) or the whole deformation (in the second case) can be removed, while retaining the local motion of intracellular structures (see Fig. 1 for an example).

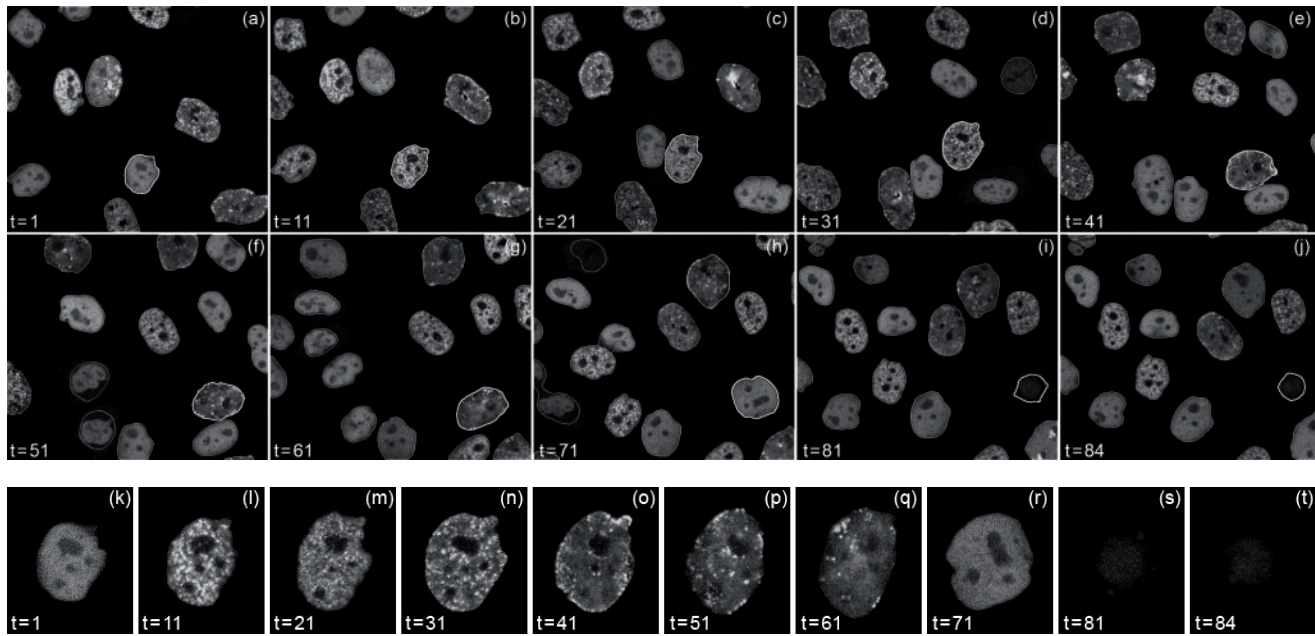


Fig. 1. Example of motion correction using the proposed approach. The two top rows show the motion of one cell extracted from a time-lapse fluorescence microscopy image dataset (outlined in white). One slice ($z = 1$) is shown for time steps 1, 11, 21, 31, 41, 51, 61, 71, 81 and 84. The third row shows (magnified) the result of cell motion correction after segmentation and tracking. In this case, only the global motion of the nucleus is subtracted.

2.2 Foci analysis

The next step after all cells are extracted from the image data is to analyze their content. For our applications, which involve the analysis of fluorescent foci, this requires two processing steps: (i) foci segmentation and (ii) foci pattern recognition.

2.2.1 Foci segmentation Similar to cell segmentation, the foci segmentation process is a challenging task, due to imperfections in the imaging and the fact that foci may vary considerably in size (from sub-resolution to large regions), as well as in local contrast, total number and degree of clustering. For example, in the case of PCNA, foci may be completely absent in both G phases of the cell cycle (Figs 1k, r–t and 2), or appear as small spots in the early-S (Figs 1l–n and 2) and middle-S (Figs 1o and 2) phases, or as a large bright blob in the late-S phase (Figs 1p, q and 2). Existing algorithms for foci segmentation were developed mostly for the analysis of FISH dots (Gué *et al.*, 2005; Kozubek *et al.*, 1999; Netten *et al.*, 1996; Raimondo *et al.*, 2005), which are easier to deal with due to their high contrast, regular spherical shape, uniform size, relatively small number, and thus relatively small degree or even complete absence of clustering. More involved methods for foci segmentation have also been proposed (Böcker and Iliakis, 2006) but these require a large number of measurements to properly handle overlapping foci regions. Here, we present a novel method for segmentation of fluorescent foci, which uses a similar ‘local’ strategy as that of Netten *et al.* (1996), but includes additional steps that also enable segmentation of heavily clustered cells of varying sizes and shapes. The segmentation pipeline consists of three steps: (i) detection of foci markers; (ii) foci segmentation; and (iii) foci selection.

In the first step of the pipeline, a marker is identified for each potential focus, to be used as seed for the actual segmentation in the second step. All local maxima of the intensity landscape are initially selected as markers. To lower the number of false positives (local maxima that do not represent actual foci) in this stage, we first perform patch-based image reconstruction (Boulanger *et al.*, 2007). Since foci may appear as relatively small structures, we use patches of size 3×3 pixels. Using larger patches may blur the

boundary between two neighboring foci so that it will become impossible to recognize them as two separate objects. Additionally, since we aim to perform the segmentation in 3D, we apply depth correction of intensity such that the mean and the variance of the intensity distribution of each slice within the cell (or nucleus) region are equal to those of the chosen reference slice. Example results after applying patch-based reconstruction and foci marker detection are shown in Figure 2b and c, respectively.

In the second step, foci segmentation is started by predicting the size of each focus, which is accomplished by calculating the average intensity in a window centered around its corresponding marker. Specifically, for each focus, the local average intensity is calculated for different sizes of the window (from unity to a pre-defined maximum value of expected foci sizes), and the estimated radius is taken to be the position of the maximum gradient of the resulting curve. In most cases, this procedure allows correct segmentation of neighboring foci, even if their sizes differ significantly, and it helps to segment large conglomerates of clustered foci (see Fig. 2 for the late-S phase). Segmentation is then performed in a window Ω_0 corresponding to the estimated size of the focus and centered around its marker, by energy minimization using the graph-cuts method (Boykov and Kolmogorov, 2003), which allows to combine both image- and smoothness-based energies:

$$E = E_{\text{image}} + E_{\text{smoothness}} \tag{3}$$

The latter is especially important in the case of noisy fluorescence microscopy images, where the boundaries of the foci are very weak. To calculate the image-based energy, we first apply the Shanbhag threshold (Shanbhag, 1994) in Ω_0 , thereby obtaining two classes: foreground (foci) and background. The image-based energy of the foreground and of the background is defined as the negative logarithm of the corresponding intensity histogram, and the total image-based energy is obtained as:

$$E_{\text{image}} = \sum_{\mathbf{x} \in \Omega_0} -\ln \frac{h_f(I(\mathbf{x}))}{h_b(I(\mathbf{x}))}, \tag{4}$$

where $I(\mathbf{x})$ is the image intensity of the voxel \mathbf{x} , and h_f and h_b are the smoothed intensity histograms of the foreground and of the background,

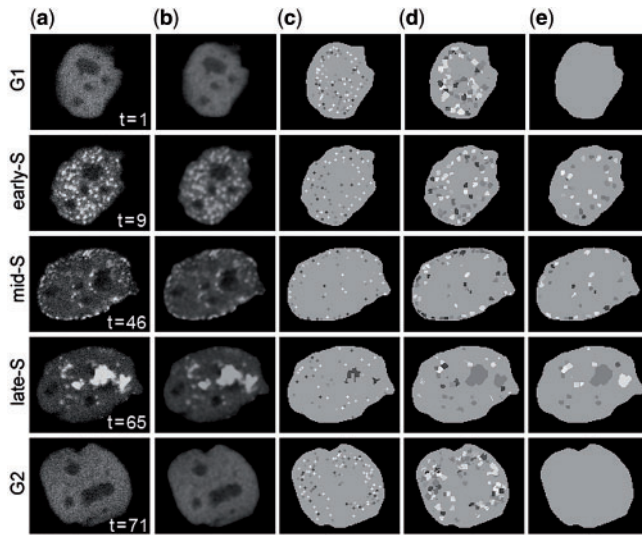


Fig. 2. Example of foci segmentation using our algorithm: (a) images of the same nucleus in five different time steps (1, 9, 46, 65, 71), each representing one of the phases of the cell cycle (G1, early-S, middle-S, late-S, G2); (b) results of applying patch-based reconstruction to each image; (c) initially detected foci markers (dots in different shades of gray); (d) results of the graph-cut-based segmentation algorithm; and (e) final results after foci selection. All images are the first slice ($z = 1$) of the corresponding 3D image stack.

respectively. The regularization energy term is defined as the sum of a certain form functional over the set N of all the neighboring voxel pairs (p, q)

$$E_{\text{smoothness}} = \lambda \sum_{(p,q) \in N} \exp\left(-\frac{\|I_p - I_q\|^2}{2\sigma^2}\right) \frac{1}{\|p - q\|}, \quad (5)$$

where λ is a real-valued weight, and the parameter σ is calculated from the data. The result of the foci segmentation step is shown in Figure 2d.

The third and final step is the selection of segmented foci in order to reject false positives. This process is guided by two parameters, which can be determined empirically: the expected minimum focus size and the expected contrast (the difference between the mean intensities of a focus region and its local background). The latter parameter is used as follows: if none of the foci whose size exceeds the minimum focus size has contrast larger than the expected contrast value, then all the foci are considered to be false. In addition, the Grubbs test (Grubbs, 1969) for detection of statistical outliers is applied. The test is performed on the intensity distribution of a local window around each segmented focus. Specifically, all voxels belonging to a segmented focus are added one-by-one to the mentioned local background distribution, and the Grubbs test is performed to detect which of these are outliers. If the number of outliers detected this way is less than the provided minimum focus size threshold, the focus is rejected. Applying all three criteria (minimum focus size, expected contrast and the statistical test) together, we obtain the final result of the foci segmentation algorithm, examples of which are shown in Figure 2e.

2.2.2 Foci pattern recognition The analysis of foci patterns is relevant to many biological studies. In this article, we consider the example of automatic identification of the cell-cycle phase based on PCNA-GFP foci. Most of the published methods for cell-cycle phase identification rely on machine learning techniques, which typically require large numbers of training data and/or the calculation of a large number of (static and dynamic) features for classification (Ersoy *et al.*, 2009; Harder *et al.*, 2006; Wang *et al.*, 2008). Alternatively, cells may be labeled explicitly with cell-phase

markers (Padfield *et al.*, 2009; Thomas, 2003), giving them a characteristic appearance during each of the cell-cycle phases. In our applications, we aim to identify the cell phases directly from the inherent labels used in the experiments. Our algorithm is based on the typical behavior of PCNA foci through each of the phases of the cell cycle. Since in this particular application we are interested in the duration of each of the phases of the cell cycle (see Section 3.2), we approach the problem by finding transitions between different phases in the complete sequence rather than trying to classify each of the images as belonging to one of the cell phases regardless of the rest of the images in the sequence. A set of simple techniques is used to distinguish the moment at which some cell goes from one stage to another.

The algorithm starts with detecting the presence of the G1 and the G2 phases by the absence of foci, keeping in mind that G1 is always the first and G2 the last phase in the sequence. Then, it detects possible transitions between the early-S and the middle-S phase, and between the middle-S and the late-S phase, using K -means clustering. Since different features are discriminative for different sub-phases of the S phase, we found it convenient to perform the clustering twice (once for the early-S and the middle-S, and once for the middle-S and the late-S phases) rather than trying to classify all three sub-phases in one step. The clustering is achieved by using only two features (time step and percentage of foci located at the boundary) in the first case, and three features (time step, percentage of foci located at the boundary and the number of foci in the upmost slice that contains foci) in the second case. Since in some of the sequences not all the phases are imaged, the missing phases should be disregarded during the cell-phase classification. In the algorithm, the decision about the existence of some of the sub-phases of S is done automatically by analyzing the range of the values of the two features: the percentage of foci located at the boundary (for the transition from the early-S to the middle-S) and the number of foci in the upmost slice that contains foci (for the transition from the middle-S to the late-S). The corresponding transition is disregarded if the maximal and the minimal value of a feature are on the same side of an empirically set threshold.

3 VALIDATION

The developed system was validated by comparing its performance to expert human observers in two experiments: (i) foci counting and (ii) foci-pattern-based cell-phase identification. For both experiments, the parameters of the system were kept fixed to the following empirically determined values: the parameters of the cell segmentation and tracking algorithm (Section 2.1.1) were set as previously described (Dzyubachyk *et al.*, 2010), the parameters of the motion compensation algorithm (Section 2.1.2) were set to $\alpha = 0.5$ and $\beta = 0.95$, the parameter λ of the smoothness energy term of the foci segmentation algorithm (Section 2.2.1) was set to 10% of the maximal value of the corresponding image-based energy, the minimal focus size (Section 2.2.1) was set to three or five voxels depending on the image size, the expected contrast (Section 2.2.1) was set to 0.05 for the first experiment (Section 3.1) and 0.2 for the second experiment (Section 3.2), and finally the thresholds used in cell-phase classification (Section 2.2.2) were set to 30% for the percentage of foci located at the boundary and five for the number of foci in the upmost slice that contains foci.

3.1 Foci counting

The protein 53BP1 forms foci in response to genotoxic stress, particularly agents inducing DNA double-strand breaks (Anderson *et al.*, 2001). Moreover, these foci are thought to represent actual sites of DNA breaks (Rodrigue *et al.*, 2006) and their disappearance is related to the DNA double-strand break repair kinetics. For example, we found in normal mouse embryonic stem (ES) cells

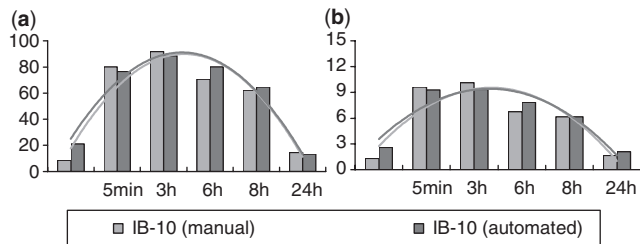


Fig. 3. Comparison between manual (light-gray) and automated (dark-gray) 53BP1 foci counting for normal ES cells (IB-10) in terms of (a) the percentage of the positive cells and (b) the average number of foci per cell at various time points. For each of the measures the corresponding values and the obtained polynomial trend lines are shown.

(IB-10) that the percentage of positive cells (containing at least 5 foci per cell) drastically increased just 5 min after treatment by 8 Gy ionizing radiation (IR) and decreased at 3 h to reach the normal situation in untreated cells at 24 h after IR (data not published; see Fig. 3).

3.1.1 Data ES cells were fixed at selected time points after IR treatment (8 Gy) and 53BP1 foci were imaged by indirect immunofluorescence using anti-53BP1 antibodies and confocal microscopy (Zeiss LSM-510) with a Plan-Apochromat 63 \times /1.4 oil-immersion objective lens. The dataset consisted of in total 49 images of size 512 \times 512 pixels (resolution 146.2 \times 146.2 μ m/pixel) or 1024 \times 1024 pixels (resolution 73.1 \times 73.1 μ m/pixel). Each image contained two channels: the DNA channel and the protein channel. The DNA channel was used for the segmentation of the cells because of its more homogeneous signal distribution in the cell regions (see Fig. 4).

3.1.2 Results The sample images in Figure 4 illustrate that the cell colonies were densely clustered and that some of the nuclei showed very irregular shapes (as imaged). Together with the relatively low and inhomogeneous contrast, and a considerable amount of noise, this makes automated segmentation challenging. Nevertheless, in all the images our system was able to yield satisfactory segmentations for subsequent foci analysis. Next, automatic foci counting was performed, and the results were compared to manual counts by an expert human observer. To make a fair comparison, for each image we selected the same number of segmented cells as considered by the human expert in the manual analysis, by applying a size threshold. In total, 858 cells were selected for automatic foci counting, 685 of which contained foci and 435 were identified as positive (≥ 5 foci). Two measures were calculated for each time point: the percentage of positive cells and the average number of foci per cell. The results (Fig. 3) clearly show that the automatically obtained results are in good agreement with the results obtained by manual analysis, both qualitatively and quantitatively. In particular, for both measures, the calculated general trend is virtually the same.

3.2 Cell-phase identification

The DNA polymerase processivity factor PCNA is central in DNA replication. We analyzed the temporal localization pattern

of GFP-tagged PCNA in living CHO cells during the different cell-cycle phases (G1, early-S, middle-S, late-S, G2). Replication of the mammalian genome starts at thousands of origins activated at different times during the S phase. By tracking the individual sites of replication foci represented by PCNA, we can investigate how this replication program is coordinated. In a previous study (Essers *et al.*, 2005b), we showed that the average times needed to progress through one complete cell cycle varies greatly in individual cells, with the largest variation in the duration of the G1 phase. Here, we aim to perform similar analysis in a fully automatic fashion using our system.

3.2.1 Data Five fluorescence microscopy image datasets were acquired as described by Essers *et al.* (2005b) using a confocal microscope (Zeiss LSM-510) with a Plan-Apochromat 63 \times /1.4 oil-immersion objective lens. The images consisted of 92 time steps (~ 10 min intervals) each having five slices (1 μ m apart) of size 512 \times 443 pixels (103.4 \times 89.5 μ m). All cell nuclei were automatically segmented, tracked and motion corrected (for retrospective visual examination; see sample results in Fig. 5), and for each of the nuclei the PCNA foci were segmented. The subsequent analysis was restricted to cells passing through at least one whole phase of the cell cycle during the time span of the sequence. In addition, cells partly falling outside the field of view at any time point were also disregarded, as these cannot be reliably analyzed due to incomplete information. This selection procedure resulted in 29 cells suitable for analysis. Two expert biologists independently marked the transition moments between the different phases of the cell cycle for each of the selected cells in the raw image data to serve as the ground truth.

3.2.2 Results The plots in Figure 6 show the differences in the phase transition times as found by our system versus both observers for each of the four possible phase transitions. The results clearly confirm, in agreement with Ersoy *et al.* (2009), that it is much easier to distinguish the G1 and G2 phases from the S phase (Figs 5 and 6a, d) than to distinguish between the different sub-phases of the S phase (Figs 5 and 6b, c). For the transitions from G1 to early-S, and from late-S to G2, the absolute differences in the times detected by our system versus any of the two observers did not exceed 3 time points, which is the same as the maximum difference found between the two observers. For the transitions from early-S and middle-S, and from middle-S to late-S, the maximum absolute difference between our system and any of the observers was 16 time points, which again is equal to the maximum difference found between the two observers. In most cases, the differences with respect to the two observers showed opposite signs (meaning that the automatically detected transition time was in between the times indicated by the observers), or one of the differences was relatively small (indicating that the automatically detected time point was close to that found by one of the observers). However, there were also several cases where our algorithm showed considerable difference with both observers, while their results were in good agreement. An important observation following from these cases is that the results were much better (closer to those of the observers) for sequences where more transitions (ideally all four) were present. And vice versa, for sequences where only two or three out of four transitions were present, our algorithm encountered difficulties in correctly detecting the time moments of those transitions.

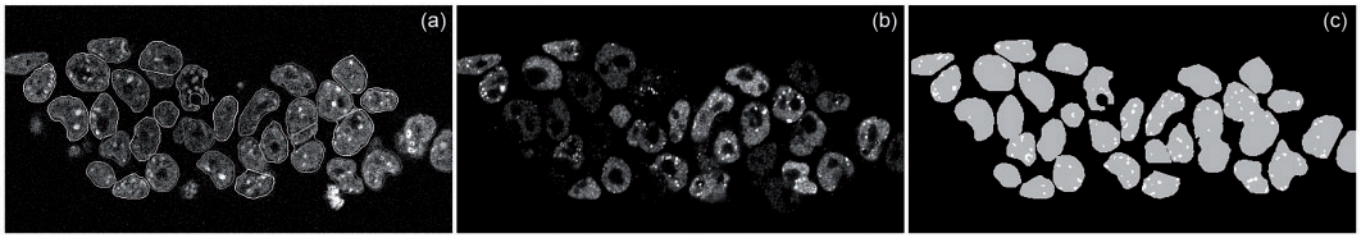


Fig. 4. Sample results from the automated foci counting experiment: (a) DNA channel with segmented cell boundaries overlaid (contours of various shades of gray); (b) protein channel; (c) region masks (gray) extracted from (a) together with the foci of interest (white) segmented from (b). Each of the images has been cropped from its original size and on the images (a) and (b) contrast enhancement was performed for better visualization.

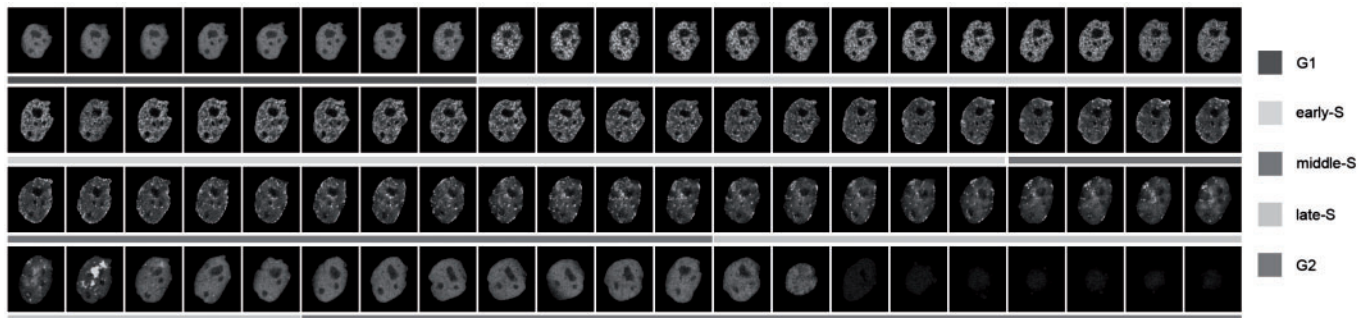


Fig. 5. Sample results from the cell-phase identification experiment. Shown from top left to bottom right are cropped images of 84 successive time points of a single, motion-corrected cell nucleus, going from the G1 phase, through the early-S, middle-S, late-S, to the G2 phase (indicated by bars in different shades of gray below the images), as automatically recognized by our system based on characteristic foci patterns for each of these phases. The example also illustrates the observation that it is easier (also visually) to distinguish the G phases from the S phases than to distinguish between the different S phases.

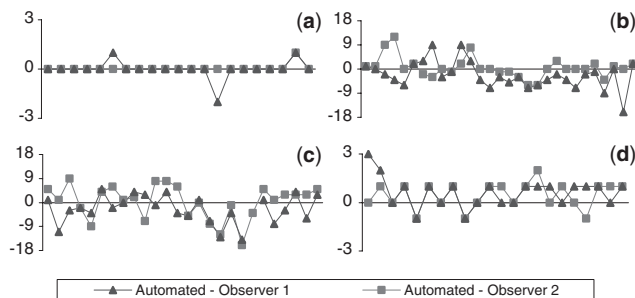


Fig. 6. Comparison between manual and automated detection of phase transition moments in PCNA-stained cells. The four plots correspond to the four possible phase transitions: (a) G1 to early-S (21 cases); (b) early-S to middle-S (29 cases); (c) middle-S to late-S (26 cases); and (d) late-S to G2 (22 cases). In each case, the difference in detection times between the automated method and each of the two observers is plotted. A missing point on one of the curves in (c) means that the corresponding phase transition was not detected by the corresponding observer.

4 DISCUSSION

In this work, we have presented our fully automated system for performing intracellular analysis at the individual-cell level. The system consists of two main parts: cell analysis (including cell segmentation, tracking, motion correction) and foci analysis

(foci segmentation and pattern analysis). The experimental results presented in this article show that the system performs comparably to manual analysis by expert biologists for the tasks of foci counting and foci-pattern-based cell-phase identification. The main contribution of the work is that the different analysis tasks are combined into an integrated, fully automated system, which does not require any user interaction (apart from inevitable initial parameter setting). An additional advantage of the system compared to some other advanced methods is that it does not involve an explicit (machine) learning stage, which would require large amounts of training data. Instead, it uses features derived directly from the segmented foci. Direct comparisons with experimental results reported in other papers on automated foci counting and cell-phase identification methods could not be made, either because these experiments focused on different applications than ours, or they were based on different quantitative measures and/or imaging protocols. However, our primary goal was to develop a system that would allow upscaling of experiments that are normally performed manually by expert human observers. Being able to reproduce their findings, our system can indeed replace tedious manual analyses, and thus enables high-content screening.

Since both motion correction and intracellular analysis rely on the results of the cell tracking, this step is of crucial importance to the overall performance of the system. The model-evolution-based tracking algorithm generally yields good results (see also Dzyubachyk *et al.*, 2010), as it better integrates temporal

information, but inherently has the tendency to propagate errors. Because of this, the initial segmentation of the first time frame is critical. Although we did not encounter serious problems in the described experiments, our method might fail to properly identify each cell in cases where cells are highly clustered, or have irregular shape and intensity (see Fig. 4). In such cases, either manual correction is needed, or more specialized (application dependent) methods incorporating additional knowledge about the problem should be used for improving the initial segmentation.

In designing this system, we have attempted to minimize the number of input parameters, and also to make them intuitive. As pointed out, the level-set-based cell segmentation and tracking algorithm plays a key role in the system, and we have previously described optimal parameter selection for this algorithm (Dzyubachyk et al., 2010). The parameters of the new components of the system were determined empirically and fixed to the values indicated in Section 3. In general, we found that the system is well behaved with respect to parameter changes, and that practical values can be easily found using only a few test images.

The current version of the system (as used in the presented experiments) was coded in MATLAB (The MathWorks, Inc., USA) and compiled for use as a stand-alone software tool. On a standard PC (Intel Pentium 4-CPU, 3.6 GHz, 3 GB RAM, running Windows XP), full cell segmentation and tracking currently takes about 3.5 h per sequence of 92 time steps, which contain 20 cells on average (i.e. about 7 s per cell, per time point), optional cell motion correction takes about 35 s per cell per time point, foci segmentation about 15 s per cell per time point, and finally the calculation of foci-related measures and cell-phase identification ~ 0.5 s in total per time point. Considerably higher speeds can be expected after conversion to a full C++ implementation and further optimization of the source code. Also, parts of the system allow a parallel implementation, which would further increase performance. This is envisaged for near-future work. The software will be made publicly available, free of charge for non-commercial use, after publication of this article.

Funding: Netherlands Organization for Scientific Research (NWO) through VIDI-grant 639.022.401 (to E.M.); European Commission through FP7-grant 201842 (the ENCITE project) (in part).

Conflict of Interest: none declared.

REFERENCES

- Al-Kofahi, O. et al. (2006) Automated cell lineage construction: a rapid method to analyze clonal development established with murine neural progenitor cells. *Cell Cycle*, **5**, 327–335.
- Anderson, L. et al. (2001) Phosphorylation and rapid relocalization of 53BP1 to nuclear foci upon DNA damage. *Mol. Cell Biol.*, **21**, 1719–1729.
- Böcker, W. and Iliakis, G. (2006) Computational methods for analysis of foci: Validation for radiation-induced γ -H2AX foci in human cells. *Radiat. Res.*, **165**, 113–124.
- Boulanger, J. et al. (2007) Space-time adaptation for patch-based image sequence restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, **29**, 1096–1102.
- Boykov, Y. and Kolmogorov, V. (2003) Computing geodesics and minimal surfaces via graph cuts. In *9th IEEE International Conference on Computer Vision (ICCV 2003)*, IEEE Computer Society, Nice, France, pp. 26–33.
- Dufour, A. et al. (2005) Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces. *IEEE Trans. Image Process.*, **14**, 1396–1410.
- Dzyubachyk, O. et al. (2010) Advanced level-set based cell tracking in time-lapse fluorescence microscopy. *IEEE Trans. Med. Imaging*, **29**, 852–867.
- Ersoy, I. et al. (2009) Segmentation and classification of cell cycle phases in fluorescence imaging. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009, 12th International Conference*, Springer, London, UK, pp. 617–624.
- Essers, J. et al. (2005a) Dynamics of relative chromosome position during the cell cycle. *Mol. Biol. Cell*, **16**, 769–775.
- Essers, J. et al. (2005b) Nuclear dynamics of PCNA in DNA replication and repair. *Mol. Cell Biol.*, **25**, 9350–9359.
- Gerlich, D. and Ellenberg, J. (2003) 4D imaging to assay complex dynamics in live specimens. *Nat. Cell Biol.*, **4**, S14–S19.
- Gordon, A. et al. (2007) Single-cell quantification of molecules and rates using open-source microscope-based cytometry. *Nat. Meth.*, **4**, 175–181.
- Grubbs, F.E. (1969) Procedures for detecting outlying observations in samples. *Technometrics*, **11**, 1–21.
- Gué, M. et al. (2005) Smart 3D-FISH: automation of distance analysis in nuclei of interphase cells by image processing. *Cytometry A*, **67**, 18–26.
- Harder, N. et al. (2006) Automated analysis of the mitotic phases of human cells in 3D fluorescence microscopy image sequences. In *Medical Image Computing and Computer-Assisted Intervention*. Springer, Berlin, pp. 840–848.
- Inagaki, A. et al. (2009) Dynamic localization of human RAD18 during the cell cycle and a functional connection with DNA double-strand break repair. *DNA Repair*, **8**, 190–201.
- Kim, I.-H. et al. (2007). Non-rigid temporal alignment of 2D and 3D multi-channel microscopy image sequences of human cells. In *Bildverarbeitung für die Medizin*, Springer, München, pp. 16–20.
- Kozubek, M. et al. (1999) High-resolution cytometry of FISH dots in interphase cell nuclei. *Cytometry A*, **36**, 279–293.
- Leonhardt, H. et al. (2000) Dynamics of DNA replication factories in living cells. *J. Cell Biol.*, **149**, 271–280.
- Li, K. et al. (2008) Cell population tracking and lineage construction with spatiotemporal context. *Med. Image Anal.*, **12**, 546–566.
- Mattes, J. et al. (2006) Analyzing motion and deformation of the cell nucleus for studying co-localizations of nuclear structures. In *Proceedings of the 2006 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, Arlington, VA, USA, pp. 1044–1047.
- Matula, P. et al. (2006) Fast point-based 3-D alignment of live cells. *IEEE Trans. Image Process.*, **15**, 2388–2396.
- Meister, P. et al. (2007) Replication foci dynamics: replication patterns are modulated by S-phase checkpoint kinases in fission yeast. *EMBO J.*, **26**, 1315–1326.
- Netten, H. et al. (1996) Fluorescent dot counting in interphase cell nuclei. *Bioimaging*, **4**, 93–106.
- Padfield, D. et al. (2009) Spatio-temporal cell cycle phase analysis using level sets and fast marching methods. *Med. Image Anal.*, **13**, 143–155.
- Paragios, N. et al. (2003) Non-rigid registration using distance functions. *Comput. Vis. Image Underst.*, **89**, 142–165.
- Raimondo, F. et al. (2005) Automated evaluation of Her-2/neu status in breast tissue from fluorescent in situ hybridization images. *IEEE Trans. Image Process.*, **14**, 1288–1299.
- Rodrigue, A. et al. (2006) Interplay between human DNA repair proteins at a unique double-strand break in vivo. *EMBO J.*, **25**, 222–231.
- Shanbhag, A.G. (1994) Utilization of information measure as a means of image thresholding. *CVGIP: Graphical Models Image Process.*, **56**, 414–419.
- Sigal, A. et al. (2006) Dynamic proteomics in individual human cells uncovers widespread cell-cycle dependence of nuclear proteins. *Nat. Meth.*, **3**, 525–531.
- Thomas, N. (2003) Lighting the circle of life: fluorescent sensors for covert surveillance of the cell cycle. *Cell Cycle*, **2**, 545–549.
- Wang, M. et al. (2008) Novel cell segmentation and online SVM for cell cycle phase identification in automated microscopy. *Bioinformatics*, **24**, 94–101.
- Yang, S. et al. (2008) Nonrigid registration of 3-D multichannel microscopy images of cell nuclei. *IEEE Trans. Image Process.*, **17**, 493–499.
- Yezzi, A.J. and Soatto, S. (2003) Deformation: deforming motion, shape average and the joint registration and approximation of structures in images. *Int. J. Comput. Vis.*, **53**, 153–167.
- Zitova, B. and Flusser, J. (2003) Image registration methods: a survey. *Image Vis. Comput.*, **21**, 977–1000.