

A Global Method for Non-Rigid Registration of Cell Nuclei in Live Cell Time-Lapse Images

Qi Gao¹ and Karl Rohr

Abstract—Non-rigid registration of cell nuclei in time-lapse microscopy images can be achieved through estimating the deformation fields using optical flow methods. In contrast to *local* optical flow models employed in the existing non-rigid registration methods, we introduce approaches based on a *global* optical flow model. Our registration model consists of a data fidelity term and a regularization term. We compared different regularizers for the deformation fields and found that a convex quadratic function is more suitable than non-convex ones. To improve the robustness, we propose an adaptive weighting scheme based on the statistics of the noise in fluorescence microscopy images as well as a combined local-global scheme. Moreover, we extend the global method by exploiting high-order image features. The best suitable high-order features are determined through learning two generative image models, namely, fields of experts and convolutional Gaussian restricted Boltzmann machine, whose model formulations are both consistent with the assumption of high-order feature constancy in the registration model. Using multiple data sets of real 2D and 3D live cell microscopy image sequences as well as synthetic image data, we demonstrate that our proposed approach outperforms the previous methods in terms of both registration accuracy and computational efficiency.

Index Terms—Non-rigid registration, optical flow, regularization, fluorescence microscopy images, restricted Boltzmann machine, Markov random field.

I. INTRODUCTION

TIME-LAPSE live cell microscopy enables studying the dynamics of subcellular structures to understand cellular processes. As the observed motion of intranuclear particles (proteins) is superimposed on the motion of cell nuclei, to analyze the relative motion patterns of particles, the particle

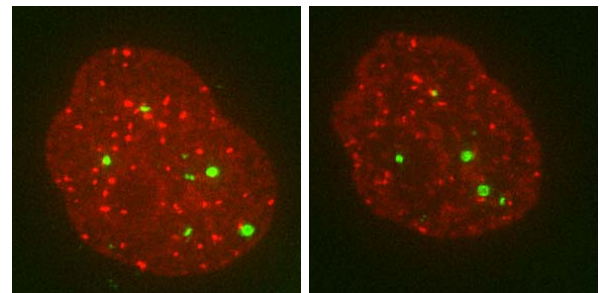


Fig. 1. Two images at different time points from a time-lapse multi-channel live cell image sequence: (left) $t = 0$; (right) $t = 50$. The red channel shows the cell nucleus and the green channel shows the intranuclear particles.

motion must be decoupled from the deformation and movement of cell nuclei. This can be achieved by registration of cell nuclei in microscopy images at different time points. Usually the intranuclear particles and cell nuclei are acquired in two different channels (see Fig. 1). The deformation fields are first estimated according to the cell nucleus channel, and then applied to the particle channel to accomplish the decoupling of the motion.

Early methods for registration of cell or cell nuclei images mostly rely on rigid and affine transformations (see [1]–[6]) and work well in applications with relatively simple conditions. As the patterns of cell deformations are usually complex and cannot be well described by linear transformations, non-rigid registration approaches have been proposed. Mattes *et al.* [7] described a method based on landmarks and thin-plate splines. Yang *et al.* [8] used segmented images and a variant of the demons algorithm with a multiresolution scheme to register static and live cell nuclei images. De Vylder *et al.* [9] proposed a registration method based on the deformation of the nuclear contour. More recently, Sorokin *et al.* [10] used contour matching and a physical prior to model the deformation based on elasticity theory. However, the approaches in [8]–[10] do not directly exploit the image intensity. Also, since the deformation fields are determined using interpolation, complex local deformations in the inner part of a cell nucleus are difficult to capture.

Another type of non-rigid cell image registration methods directly exploits the image intensity and employs *local* optical flow models to compute the deformation. These methods are based on the observation that the deformation between two consecutive image frames changes gradually over time and

Manuscript received December 22, 2018; revised February 14, 2019; accepted February 17, 2019. Date of publication February 27, 2019; date of current version October 1, 2019. This work was supported by the DFG Cluster of Excellence CellNetworks under Grant EXC 81, by the Math-Clinic Core Facility at Heidelberg University, and by the BMBF Project CancerTelSys through the e:Med Program under Grant 01ZX1602. (Corresponding author: Qi Gao.)

Q. Gao is with CellNetworks, Heidelberg University, 69120 Heidelberg, Germany, and also with the Biomedical Computer Vision Group, BioQuant, IPMB, Heidelberg University, 69120 Heidelberg, Germany.

K. Rohr is with the Biomedical Computer Vision Group, BioQuant, IPMB, Heidelberg University, 69120 Heidelberg, Germany, and also with the German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany.

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2019.2901918

shows temporal coherence. Kim *et al.* [11] extended the Lucas-Kanade optical flow model [12] to estimate dense deformation fields between consecutive frames and then register each frame to the reference frame using an incremental scheme. This method was extended by Tektonidis *et al.* [13], where a multi-frame approach was proposed to improve the robustness to image noise. Tektonidis and Rohr [14] introduced a diffeomorphic multi-frame approach, which guarantees invertibility and continuity of the computed transformations. The major advantage of optical flow-based registration methods is that they can cope with general motion which is not necessarily constrained by a physical prior. All above mentioned approaches for non-rigid registration of cell nuclei using optical flow models [11], [13], [14] are *local*, since the deformation vectors are computed based on local image patches. Besides low efficiency, using local patches implies that only some basic regularization on the result is included.

In contrast, *global* optical flow models explicitly integrate the optical flow constraint (*e.g.*, brightness constancy) and the regularization of optical flow vectors into one single objective function. The optical flow vectors for all pixels, *i.e.*, a dense optical flow field, are computed simultaneously and efficiently based on the whole images. Interpolation as often used in local models is not required. By appropriately choosing the type of regularization, global models can generally improve the performance compared to local models. For a recent survey on optical flow models, readers are refer to [15]. However, popular global optical flow models have been originally proposed for video images of natural scenes (*e.g.*, street traffic). For cell microscopy images, the image structures are generally more blurry (*e.g.*, out-of-focus local regions, scattering of fluorescence) and the image noise is more significant. Another characteristic is that occlusions (due to 3D-2D projection) generally do not occur in contrast to video images of natural scenes. Due to different properties, direct application of these global optical flow models for estimating the deformation fields in fluorescence microscopy images generally does not yield optimal results.

In addition, previous optical flow-based registration methods for cell microscopy images [6], [11], [13], [14] assume brightness constancy. This means that the image intensities of moving structures do not change over time, which is a relatively strong assumption and often violated in real data. In the field of optical flow estimation for images of natural scenes, extensions of brightness constancy to gradient constancy and high-order feature constancy were proposed to improve, for example, the robustness to varying illumination (see [16]). While it is often hard to handcraft the best suitable features, they could be determined by some feature learning method. In [16], estimation of optical flow for images of natural scenes relied on discriminatively learned features through maximizing the conditional likelihood of the data fidelity term. However, for cell microscopy image data, due to the lack of ground truth of deformations, an unsupervised feature learning method should be employed. Note that, since the feature-constancy data term in an optical flow model implies (conditional) independence among features, an ideal feature learning model should be “coupled” with the registration model (data term),

i.e., their hidden assumptions should be consistent. Unfortunately, common unsupervised feature learning methods (*e.g.*, PCA, autoencoder, sparse coding [17]) and patch-based image models with features learned simultaneously (*e.g.*, product of Student-t model [18]) do not fulfill this requirement.

For registration of temporal medical images (*e.g.*, CT, MR and US images of organs), group-wise approaches have been proposed (see [19]–[23]), which simultaneously exploit all frames of an image sequence. Typically periodic movement is considered (*e.g.*, lung, heart) and often parametric registration approaches are employed (*e.g.*, using B-splines). For registration of live cell microscopy image sequences, it was shown that optical flow-based methods with an incremental scheme [13], [14] perform better than a group-wise method [19]. A main reason is that an incremental scheme better copes with the heterogeneous changes of the intensity structure over time, since the deformations are computed by concatenating deformation fields estimated from a limited number of consecutive frames.

In this paper, we introduce *global* optical flow models with appropriate regularization of the deformation fields for non-rigid registration of cell nuclei in live-cell fluorescence microscopy images using an incremental scheme. We compare different regularizers and show that a convex quadratic function is more suitable than non-convex functions widely used for optical flow estimation in video images of natural scenes. To increase the robustness to noise, we propose an adaptive weighting scheme derived from the mixed Poisson-Gaussian statistics of typical noise in microscopy images. The deformation fields can be efficiently computed by solving linear systems of equations.

Moreover, we extend the global models by learned image features. We show that two generative image models are consistent with the hidden assumption in the feature-constancy data term of the registration model: The first generative model is a variant of Fields of Experts (FoE) [24], [25], which was previously used for image deconvolution and denoising (see [26]–[28]). The second model is a Gaussian restricted Boltzmann machine (RBM) [29] with full-convolutional weight sharing, denoted as convolutional Gaussian RBM (cGRBM), which has been used for feature learning for natural and medical images (see [30], [31]). An additional advantage of using these two models is that they only require moderate amount of training data due to the convolutional model structure. In our approach, we address the problem that learned coefficients in the filters might shift and distort image correspondences in the feature space (note that this is not an issue in image deconvolution and denoising applications). To cope with this, we augment the standard learning procedure [25], [29] by an additional constraint. The learned FoE filters and cGRBM filters are used in our proposed global models. In addition, our adaptive weighting scheme based on brightness constancy can also be extended to the global model using high-order features to increase the robustness to noise. To deal with outliers and complex noise not following Poisson-Gaussian statistics, we also employ a combined local-global (CLG) scheme [32].

To evaluate the performance of our proposed global methods, we use multiple data sets including both 2D and 3D real

live cell microscopy image sequences as well as synthetic image data. Experimental results demonstrate that our global method significantly improves the registration accuracy compared to existing state-of-the-art non-rigid registration methods such as [10] and [14]. It also turns out that our proposed method based on global optical flow models is much more efficient than existing methods based on local models. This work extends our previous conference article [33], where learned features as well as a combined local-global scheme were not used. Additionally, we have performed a much more comprehensive performance evaluation. Also, we present results for 3D live-cell microscopy images and for synthetic image data.

The paper is organized as follows: In the next section, we introduce our registration method based on global optical flow models using image intensity. Then, we extend the global model using high-order features and describe the training procedure. Afterwards, experimental results for real live cell images and for synthetic image data are presented and discussed. Finally, we give a summary.

II. GLOBAL REGISTRATION METHOD

Our non-rigid registration method for live cell microscopy images is based on a global optical flow model for estimating the deformations between consecutive image frames and uses an incremental scheme to register a temporal image sequence.

A. Incremental Temporal Registration

Given temporal cell microscopy image sequences, the goal is to determine the dense deformation field $\mathbf{w}^{0,t}$ to register the frame \mathbf{x}^t at time point t to the reference frame \mathbf{x}^0 at the first time point. However, directly estimating all $\mathbf{w}^{0,t}$, $t = 1, \dots, T$, is not a good choice, since the deformations over a long time interval are generally too complex. Instead, we estimate all deformation fields $\mathbf{w}^{t-1,t}$ between pairs of consecutive image frames and then concatenate them incrementally to obtain

$$\mathbf{w}^{0,t} = \mathbf{w}^{t-1,t} \circ \mathbf{w}^{0,t-1} = \mathbf{w}^{t-1,t} + \tilde{\mathbf{w}}^{0,t-1}, \quad (1)$$

where $\tilde{\mathbf{w}}^{0,t-1}$ is the result of using $\mathbf{w}^{t-1,t}$ to warp $\mathbf{w}^{0,t-1}$ using bilinear interpolation, and $\mathbf{w}^{0,0} \triangleq \mathbf{0}$. Below, we describe methods based on a global optical flow model for computing a deformation field $\mathbf{w}^{t-1,t}$ from two consecutive frames. The superscript of \mathbf{w} will be dropped for easier readability.

B. Global Optical Flow-Based Models

The deformation field \mathbf{w} between two consecutive images $\mathbf{x}^0 \in \mathbb{R}^N$ and $\mathbf{x}^1 \in \mathbb{R}^N$ (N is the number of pixels in an image frame) is modeled via the posterior density in a Bayesian framework

$$p(\mathbf{w}|\mathbf{x}^0, \mathbf{x}^1) \propto p_D(\mathbf{x}^0|\mathbf{w}, \mathbf{x}^1) \cdot p_R(\mathbf{w}), \quad (2)$$

where the data-dependent likelihood $p_D(\mathbf{x}^0|\mathbf{w}, \mathbf{x}^1)$ models how \mathbf{x}^0 is generated from \mathbf{w} and \mathbf{x}^1 , and the prior $p_R(\mathbf{w})$ imposes regularization on \mathbf{w} . For 2D image sequence, $\mathbf{w} \in \mathbb{R}^{2N}$ consists of x - and y -components \mathbf{u} and \mathbf{v} .

1) Data Term: When assuming brightness constancy, the data fidelity (likelihood) term can be written as

$$p_D(\mathbf{x}^0|\mathbf{w}, \mathbf{x}^1) \propto \prod_{i=1}^N \exp\left\{-\frac{1}{\alpha_i} \cdot \phi_D(\mathbf{x}_i^0 - \mathbf{x}_{\mathbf{w}i}^1)\right\}, \quad (3)$$

where α_i is the regularization weight and $\mathbf{x}_{\mathbf{w}i}^1$ is the warped (deformed) image of \mathbf{x}^1 towards \mathbf{x}^0 by the deformation field \mathbf{w} to be estimated. The penalty function ϕ_D can take an arbitrary form. When a *convex* quadratic function $\phi_D(x) = x^2$ is used, (3) will be a Gaussian function, meaning that the difference $\mathbf{x}_i^0 - \mathbf{x}_{\mathbf{w}i}^1$ is assumed to follow a Gaussian distribution. In popular optical flow methods for video images of natural scenes (see [16]), *non-convex* functions for ϕ_D are used (*e.g.*, a Lorentzian function). The reason is that, due to, for example, occlusion in video images of natural scenes, the brightness constancy assumption is violated, leading to heavy-tailed histograms with a sharp peak of the intensity differences. In contrast, for cell microscopy images, there exist less violations of the brightness constancy assumption (*e.g.*, occlusions due to 3D-2D projection as in video images generally do not occur), but significant noise. Note that for images with strong noise (*e.g.*, microscopy images), the histogram of intensity differences has a more rounded peak [16]. Therefore, the histogram of intensity differences for cell microscopy images is expected to have a shape that is largely similar to a Gaussian function. Thus, a convex quadratic function for ϕ_D should be well suited, and we use it in our method. In addition, a quadratic data term eases the inference procedure.

2) Regularization Term: For the regularization term (prior), we use a Markov random field (MRF) formulation

$$p_R(\mathbf{w}) \propto \prod_{r=1}^2 \prod_{i=1}^N \exp\left\{-\phi_R((\mathbf{d}_r * \mathbf{u})_i) - \phi_R((\mathbf{d}_r * \mathbf{v})_i)\right\}, \quad (4)$$

where \mathbf{d}_r are first-order derivative filters in x - and y -directions. The choice of the regularization function ϕ_R in (4) should depend on the statistical properties of \mathbf{w} . Unfortunately, for cell microscopy images, due to the lack of the ground truth, the statistical properties of \mathbf{w} are not known. Tseng *et al.* [34] found that the cell nucleus is stiffer and more elastic than the cytoplasm. We accordingly assume that the deformation vectors in the nuclei change gradually in space and that strong discontinuities (*e.g.*, edges) in \mathbf{w} are unlikely, which means that the filter responses will hardly have large values and their histograms will not show a heavy-tailed shape. Thus, it is expected that a convex quadratic regularization $\phi_R(x) = x^2$, which has no edge-preserving property, is more appropriate. As a comparison, we also consider two robust regularizers widely used in popular optical flow models for video images of natural scenes: the Charbonnier function $\phi_R(x) = \sqrt{x^2 + \epsilon^2}$ [32], which is a differentiable variant of the $L1$ norm, and the non-convex Lorentzian function $\phi_R(x) = \log(1 + \frac{x^2}{2\sigma^2})$ [35]. The latter arises in the context of a Student t -distribution for one degree of freedom, and is also a common potential function or regularizer in image modeling and analysis (see [36], [37]).

3) Inference: The deformation field \mathbf{w} is computed by minimizing the model energy $E(\mathbf{w}) = -\log p(\mathbf{w}|\mathbf{x}^o, \mathbf{x}^1)$, which is equivalent to the maximum a-posteriori (MAP) estimate. For small deformations, the data term can be linearized using a first order Taylor expansion. When both ϕ_D and ϕ_R are quadratic functions and a single regularization weight $\alpha_i = \alpha_0$ is applied, the model energy is

$$E(\mathbf{w}) = \|\mathbf{\Gamma}\mathbf{u} + \mathbf{\Lambda}\mathbf{v} + \mathbf{x}^1 - \mathbf{x}^o\|^2 + \alpha_0 \cdot \sum_r \sum_i ((\mathbf{d}_{ir}^T \mathbf{u})^2 + (\mathbf{d}_{ir}^T \mathbf{v})^2) + c, \quad (5)$$

where $\mathbf{\Gamma} = \text{diag}(\mathbf{x}_x^1)$ and $\mathbf{\Lambda} = \text{diag}(\mathbf{x}_y^1)$ are diagonal matrices containing the partial image derivatives \mathbf{x}_x^1 and \mathbf{x}_y^1 in x- and y-directions, \mathbf{d}_{ir} are vectors defined as $\mathbf{d}_{ir}^T \mathbf{u} = (\mathbf{d}_r * \mathbf{u})_i$, and c is a constant. To minimize (5), we need to solve the linear system of equations

$$\left(\begin{bmatrix} \mathbf{\Gamma}\mathbf{\Gamma} & \mathbf{\Gamma}\mathbf{\Lambda} \\ \mathbf{\Gamma}\mathbf{\Lambda} & \mathbf{\Lambda}\mathbf{\Lambda} \end{bmatrix} + \alpha_0 \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \right) \cdot \mathbf{w} = \begin{bmatrix} \mathbf{\Gamma} \\ \mathbf{\Lambda} \end{bmatrix} \cdot (\mathbf{x}^o - \mathbf{x}^1), \quad (6)$$

where

$$\mathbf{D} = \sum_r \sum_i \mathbf{d}_{ir} \mathbf{d}_{ir}^T.$$

As the coefficient matrix on the left side (comprising $\mathbf{\Gamma}$, $\mathbf{\Lambda}$, and \mathbf{D}) is sparse and positive definite, (6) can be efficiently solved using Cholesky factorization. For the Lorentzian and Charbonnier functions, we follow [38] and use a graduated non-convexity (GNC) scheme.

C. Handling Noise and Outliers

To cope with significant noise and outliers in cell microscopy images, we use an adaptive weighting scheme and a combined local-global scheme as well as a combination of these two schemes.

1) Adaptive Weighting (AW) Scheme: The noise in cell microscopy images largely follows mixed Poisson-Gaussian statistics. An observed image can be written as $\mathbf{x}_i = g_0 \mathbf{t}_i + \mathbf{n}_i$, where g_0 is the gain of the overall electronic system, $\mathbf{t}_i \in \mathbb{Z}_0^+$ is the number of collected photo-electrons at pixel i which is a random variable following a Poisson distribution $\mathbf{t}_i \sim \text{Poi}(\lambda_i)$, $\lambda_i \in \mathbb{R}^+$, and $\mathbf{n}_i \in \mathbb{R}$ is Gaussian noise $\mathbf{n}_i \sim \mathcal{N}(\mu_n, \sigma_n^2)$ (see [39]). The corresponding noise-free image is

$$\hat{\mathbf{x}}_i = g_0 \lambda_i + \mu_n. \quad (7)$$

Assuming brightness constancy for the (hidden) noise-free images $\hat{\mathbf{x}}_i^o = \hat{\mathbf{x}}_{\mathbf{w}i}^1$, we have $\lambda_i^o = \lambda_{\mathbf{w}i}^1$, thus $\mathbf{t}_{\mathbf{w}i}^1 \sim \text{Poi}(\lambda_{\mathbf{w}i}^1) = \text{Poi}(\lambda_i^o)$ following the same Poisson distribution as $\mathbf{t}_i^o \sim \text{Poi}(\lambda_i^o)$, and therefore the distribution of their difference can be approximated by a Gaussian [40]

$$(\mathbf{t}_i^o - \mathbf{t}_{\mathbf{w}i}^1) \sim \mathcal{N}(0, 2\lambda_i^o). \quad (8)$$

With $(\mathbf{n}_i^o - \mathbf{n}_{\mathbf{w}i}^1) \sim \mathcal{N}(0, 2\sigma_n^2)$, the intensity difference of the observed images in the *data term* (3)

$$\mathbf{x}_i^o - \mathbf{x}_{\mathbf{w}i}^1 = g_0(\mathbf{t}_i^o - \mathbf{t}_{\mathbf{w}i}^1) + (\mathbf{n}_i^o - \mathbf{n}_{\mathbf{w}i}^1) \quad (9)$$

has a Gaussian distribution with varying variance

$$(\mathbf{x}_i^o - \mathbf{x}_{\mathbf{w}i}^1) \sim \mathcal{N}(0, 2g_0^2 \lambda_i^o + 2\sigma_n^2) \propto \exp \left\{ -\frac{1}{4(g_0^2 \lambda_i^o + \sigma_n^2)} \cdot (\mathbf{x}_i^o - \mathbf{x}_{\mathbf{w}i}^1)^2 \right\}. \quad (10)$$

Comparing (10) and (3) suggests using a quadratic penalty in the data term as well as a varying regularization weight at each pixel location

$$\alpha_i = \alpha_0(g_0^2 \lambda_i^o + \sigma_n^2). \quad (11)$$

In practice, however, estimating the parameters in (11) is difficult and an approximation is needed.

We now consider the *regularization term*. In our fluorescence microscopy images, cell nuclei have more or less homogeneous intensity and include bright spot-like structures. The spots generally preserve their sizes and do not expand or shrink. Thus, it is reasonable to assume that the deformation vectors of the pixels inside one spot are very similar. Such deformation continuity can be achieved by using a higher weight for the regularization term to more strongly penalize the variation of the deformation. Since pixels inside the spots have high intensities, and higher weights are needed at these locations, we suggest using weights that are proportional to the image intensities:

$$\alpha_i = \alpha_0 \hat{\mathbf{x}}_i^o = \alpha_0(g_0 \lambda_i^o + \mu_n^o). \quad (12)$$

Thus, the weight α_i should be determined so that it is consistent with the weighting schemes for the data term in (11) and the regularization term in (12). We notice that the relations in (11) and (12) have the same structure: In both cases the weights are linear functions of λ_i and have the same sign of derivatives (simultaneously increasing or decreasing). Therefore, we propose determining the weights using the relation in (12), where the noise-free image is roughly approximated by applying Gaussian smoothing on the original image \mathbf{x}^o

$$\alpha_i = \alpha_0(\hat{\mathbf{x}}_i^o + b) \approx \alpha_0((\mathbf{g}_\sigma * \mathbf{x}^o)_i + b), \quad (13)$$

where b is a positive constant and \mathbf{g}_σ is a Gaussian filter with kernel width σ .

For the proposed adaptive weighting scheme, the model energy is still convex. The deformation field is computed by solving

$$\left(\begin{bmatrix} \mathbf{\Gamma}\mathbf{\Gamma} & \mathbf{\Gamma}\mathbf{\Lambda} \\ \mathbf{\Gamma}\mathbf{\Lambda} & \mathbf{\Lambda}\mathbf{\Lambda} \end{bmatrix} + \begin{bmatrix} \mathbf{D}_\alpha & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_\alpha \end{bmatrix} \right) \cdot \mathbf{w} = \begin{bmatrix} \mathbf{\Gamma} \\ \mathbf{\Lambda} \end{bmatrix} \cdot (\mathbf{x}^o - \mathbf{x}^1), \quad (14)$$

where $\mathbf{D}_\alpha = \sum_{i,r} \alpha_i \mathbf{d}_{ir} \mathbf{d}_{ir}^T$.

2) Combined Local-Global (CLG) Scheme: The statistics of noise and outliers in cell microscopy images are sometimes more complex and cannot be accurately modeled using mixed Poisson-Gaussian as in the case of the adaptive weighting scheme. An alternative to improve the robustness to noise and outliers is the combined local-global (CLG) method [32] which exploits local image information within a global model.

The deformation field \mathbf{w} is then determined by solving

$$\begin{pmatrix} \text{diag}\{\mathbf{g}_\rho * (\mathbf{x}_x^1 \odot \mathbf{x}_x^1)\} & \text{diag}\{\mathbf{g}_\rho * (\mathbf{x}_x^1 \odot \mathbf{x}_y^1)\} \\ \text{diag}\{\mathbf{g}_\rho * (\mathbf{x}_x^1 \odot \mathbf{x}_y^1)\} & \text{diag}\{\mathbf{g}_\rho * (\mathbf{x}_y^1 \odot \mathbf{x}_y^1)\} \end{pmatrix} + \alpha_0 \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \mathbf{w} = \begin{bmatrix} \text{vec}\{\mathbf{g}_\rho * (\mathbf{x}_x^1 \odot (\mathbf{x}^0 - \mathbf{x}^1))\} \\ \text{vec}\{\mathbf{g}_\rho * (\mathbf{x}_y^1 \odot (\mathbf{x}^0 - \mathbf{x}^1))\} \end{bmatrix}, \quad (15)$$

where “ \odot ” denotes element-wise product and \mathbf{g}_ρ is a Gaussian kernel with standard deviation ρ . Note that, as the sparsity of the coefficient matrix in (15) is the same as in (6), the solution can still be computed efficiently.

3) Combined AW + CLG Scheme: The AW and CLG schemes described above can also be combined. To this end we extended the linear system of equations of the CLG scheme in (15) by using the weight matrix \mathbf{D}_α from the AW scheme in (14) instead of the matrix \mathbf{D} .

D. Extension to 3D Image Data

For temporal 3D image sequences (volume images over time), the deformation field $\mathbf{w} \in \mathbb{R}^{3N}$ includes an additional z-component $\mathbf{h} \in \mathbb{R}^N$. While the formulation of the data term (3) is the same as in the 2D case, the regularization term needs to be extended to regularize all components of \mathbf{w}

$$p_R(\mathbf{w}) \propto \prod_{r=1}^3 \prod_{i=1}^N \exp\{-\phi_r((\mathbf{d}_r * \mathbf{u})_i) - \phi_r((\mathbf{d}_r * \mathbf{v})_i) - \phi_r((\mathbf{d}_r * \mathbf{h})_i)\}, \quad (16)$$

where \mathbf{d}_3 is a first-order derivative filter in z-direction. The linear system of equations for computing the 3D deformation field \mathbf{w} for a global model (cf. (5)) is

$$\begin{pmatrix} \begin{bmatrix} \Gamma\Gamma & \Gamma\Lambda & \Gamma\Delta \\ \Gamma\Lambda & \Lambda\Lambda & \Lambda\Delta \\ \Gamma\Delta & \Lambda\Delta & \Delta\Delta \end{bmatrix} + \alpha_0 \begin{bmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D} \end{bmatrix} \end{pmatrix} \mathbf{w} = \begin{bmatrix} \Gamma \\ \Lambda \\ \Delta \end{bmatrix} (\mathbf{x}^0 - \mathbf{x}^1), \quad (17)$$

where $\Delta = \text{diag}(\mathbf{x}_z^1)$ with \mathbf{x}_z denoting the partial derivatives in z-direction. Γ , Λ , and \mathbf{D} are defined analogously as in the 2D case described above. Note that, as the dimension of the coefficient matrix in (17) is much higher than that for the 2D model, using Cholesky factorization to compute the exact solution is not efficient. Thus, for the 3D model we instead use the preconditioned conjugate gradient (PCG) method [41] to obtain an approximate solution. To handle noise and outliers, the adaptive weighting or the CLG scheme (see Sec. II-C above) can be straightforwardly applied.

III. REGISTRATION IN LEARNED FEATURE SPACES

For estimating the optical flow or deformation field between two images \mathbf{x}^0 and \mathbf{x}^1 , often brightness constancy is assumed and the brightness residual after image warping $\delta_i = \mathbf{x}_i^0 - \mathbf{x}_{\mathbf{w}i}^1$ is penalized in the data energy term $E_D(\mathbf{w}) = \sum_i \xi(\delta_i)$, where ξ is some penalty function (cf. Sec. II-B.1). To improve the performance under varying illumination, gradient constancy and high-order feature constancy were proposed for images of

natural scenes (see [16]). Suppose \mathbf{f}_k are filters for extracting image derivatives or high-order features, the residual is then $\delta_{ki} = (\mathbf{f}_k * \mathbf{x}^0)_i - (\mathbf{f}_k * \mathbf{x}^1)_{\mathbf{w}i}$. When several features are jointly considered, all energy terms of the features are summed up yielding the data term $E_D(\mathbf{w}) = \sum_k \sum_i \xi(\delta_{ki})$, whose corresponding probabilistic formulation (the likelihood) is

$$\begin{aligned} p_D(\mathbf{w}) &\propto \exp\{-E_D(\mathbf{w})\} \\ &\propto \prod_{k=1}^K \prod_{i=1}^N \exp\{-\xi(\delta_{ki})\} \\ &\propto \prod_{k=1}^K \prod_{i=1}^N p((\mathbf{f}_k * \mathbf{x}^0)_i | \mathbf{w}, (\mathbf{f}_k * \mathbf{x}^1)_i). \end{aligned} \quad (18)$$

The factorization of the likelihood into a product of conditional distributions $p((\mathbf{f}_k * \mathbf{x}^0)_i | \mathbf{w}, (\mathbf{f}_k * \mathbf{x}^1)_i)$ implies that the filter responses $(\mathbf{f}_k * \mathbf{x}^0)_i$ are assumed to be (conditionally) independent. However, in practice the requirement of independence cannot be fulfilled, as there are more filter responses than pixels. Note that orthogonality of filters does not lead to independence of filter responses (e.g., image derivatives in x- and y-directions are generally correlated).

A. Learning Features Using Generative Image Models

According to (18), high-order features with least dependence are best. In addition, the features should be specific for the considered images. To determine filters that capture the most important and least dependent features, we train filter-based generative image models implying the independence assumption. We consider two generative image models: Fields of Experts (FoE) and convolutional Gaussian restricted Boltzmann machine (cGRBM). These two kinds of models not only have been widely used in computer vision, but also have the advantage that they only require moderate amount of training data due to the convolutional model structure.

1) Fields of Experts: The FoE [24], [25] is a filter-based, high-order Markov random field (MRF) image prior, where both potential functions and filters are trained unsupervisedly. The probability density of an image \mathbf{x} is defined as

$$p_{\text{FoE}}(\mathbf{x}) \propto \prod_{k=1}^K \prod_{i=1}^N \psi((\mathbf{f}_k * \mathbf{x})_i; \omega_k), \quad (19)$$

where \mathbf{f}_k are linear filters, and $\psi(\cdot; \omega_k)$ are filter-specific potential functions (or experts) with parameters ω_k .

Comparing (19) with (18), it can be seen that the FoE prior is a generalization of the data term (18). Proper training procedures will not only yield filters that capture specific image features of the training data, but also enforce the filter responses to be as independent as possible, making the learned filters well suitable for the model in (18).

2) Convolutional Gaussian RBM: The cGRBM is a variant of a convolutional RBM [42], [43], for which a convolutional weight sharing scheme (as for convolutional neural networks) is used for a Gaussian RBM [29]. The shared weights can be regarded as filters. In a cGRBM, an image is modeled by defining an energy function of real-valued visible units \mathbf{x} (here, the image pixels) and binary hidden units \mathbf{z} :

$$E_{\text{cGRBM}}(\mathbf{x}, \mathbf{z}) = \frac{1}{2} \sum_{i=1}^N \mathbf{x}_i^2 - \sum_{k=1}^K \sum_{i=1}^N z_{ki} ((\mathbf{f}_k * \mathbf{x})_i + b_k), \quad (20)$$

where \mathbf{f}_k are the weights or filters, and b_k denote the biases.

The probability density of \mathbf{x} can be obtained by marginalizing out the hidden units

$$\begin{aligned} p_{\text{cGRBM}}(\mathbf{x}) &\propto \mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{I}) \cdot \prod_{k=1}^K \prod_{i=1}^N (1 + \exp\{(\mathbf{f}_k * \mathbf{x})_i + b_k\}) \\ &\propto \mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{I}) \cdot \prod_{k=1}^K \prod_{i=1}^N \varphi((\mathbf{f}_k * \mathbf{x})_i; b_k), \end{aligned} \quad (21)$$

where φ are filter specific potential functions. From (21), it can be seen that the cGRBM is also consistent with the independence property of model in (18).

3) Learning: Although we are only interested in the coefficients of the filters \mathbf{f}_k , all model parameters (including ω_k for the FoE and b_k for the cGRBM) must be learned simultaneously. As the partition functions of both models are intractable, learning is non-trivial. In our approach we use sampling-based approximate maximum likelihood learning and follow the best practices suggested in [25] and [29]. We noticed that these learning procedures do not constrain the learned coefficients from shifting within the filters, which is not an issue in other applications (*e.g.*, image restoration [26]–[28]). However, shifting of the filter coefficients generally leads to shifted extracted image features (see Fig. 2), which may distort image correspondences. To address this, we include an additional constraint in the learning procedure. In our approach, after each iteration of the filter update, the filter coefficients are shifted so that their gravity center is always the center of the filter, and thus the extracted features are not shifted.

B. Non-Rigid Registration Using Learned Features

Based on the learned filters \mathbf{f}_k using FoE or cGRBM, we define the data term assuming feature constancy

$$p_D(\mathbf{x}^0 | \mathbf{w}, \mathbf{x}^1) \propto \prod_{k=1}^K \prod_{i=1}^N \exp\left\{-\frac{1}{\alpha_{ki}} ((\tilde{\mathbf{x}}_k^0)_i - (\tilde{\mathbf{x}}_{k\mathbf{w}}^1)_i)^2\right\}, \quad (22)$$

where $\tilde{\mathbf{x}}_k \triangleq \mathbf{f}_k * \mathbf{x}$ and $\tilde{\mathbf{x}}_{k\mathbf{w}}$ denotes the warped (high-order feature) image by \mathbf{w} . We use a quadratic penalty function according to the analysis in Sec. II-B above. If we use a single regularization weight $\alpha_{ki} = \alpha_0$ in (22), \mathbf{w} can be computed by solving the linear system of equations

$$\begin{aligned} \left(\alpha_0 \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} + \sum_{k=1}^K \begin{bmatrix} \mathbf{\Gamma}_k \mathbf{\Gamma}_k & \mathbf{\Gamma}_k \mathbf{\Lambda}_k \\ \mathbf{\Gamma}_k \mathbf{\Lambda}_k & \mathbf{\Lambda}_k \mathbf{\Lambda}_k \end{bmatrix} \right) \cdot \mathbf{w} \\ = \sum_{k=1}^K \begin{bmatrix} \mathbf{\Gamma}_k \\ \mathbf{\Lambda}_k \end{bmatrix} \cdot (\tilde{\mathbf{x}}_k^0 - \tilde{\mathbf{x}}_k^1), \end{aligned} \quad (23)$$

where the diagonal matrices $\mathbf{\Gamma}_k = \text{diag}\{\tilde{\mathbf{x}}_{kx}^1\}$ and $\mathbf{\Lambda}_k = \text{diag}\{\tilde{\mathbf{x}}_{ky}^1\}$ are composed of the partial derivatives in x - and y -directions of $\tilde{\mathbf{x}}_k^1$. As the sparsity of the coefficient matrix is the same as for the global model using brightness constancy in (6), solving (23) is still very efficient.

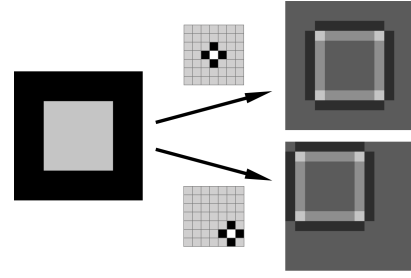


Fig. 2. Application of two filters (middle) to an input image (left). The result (right) shows that the extracted features are shifted using the filter with shifted coefficients (middle-bottom).

TABLE I
REAL TIME-LAPSE MICROSCOPY IMAGES FOR EVALUATION

Sequence	Length (time points)	Dimension (pixel)	Pixel/voxel size (nm)
S1	150	512 × 512	216 × 216
S2	200	384 × 384	216 × 216
S3	100	512 × 512	216 × 216
S4	149	512 × 512	216 × 216
T1	100	512 × 512 × 15	216 × 216 × 500
T2	149	512 × 512 × 10	216 × 216 × 1500
A1	25	512 × 512	240.5 × 240.5
A2	38	512 × 512	240.5 × 240.5
B1	42	287 × 356	470 × 470
B2	30	512 × 512	490 × 490
B3	42	279 × 318	540 × 540
B4	35	322 × 304	540 × 540

Constancy assumptions based on high-order image features are more sensitive to noise. Although Poisson-Gaussian distributed noise is more complex in feature space (compared to brightness), due to the linearity of learned filters, the adaptive weighting scheme for high-order features can be derived analogously as for brightness described in Sec. II-C.1 above. In addition, the combined local-global (CLG) scheme described in Sec. II-C.2 can also be used.

IV. EXPERIMENTS

A. Data

We use multiple data sets of real time-lapse fluorescence microscopy images of live cells to evaluate our proposed methods (see Tab. I). Data set “S” consists of four 2D image sequences (S1–S4) and data set “T” contains two 3D volume image sequences (T1 and T2). All these sequences were acquired by a widefield microscope and consist of two channels, one for nuclei of live human cells (U2OS cell line) with different chromatin stainings (H2A-mCherry, YFP-SP100) and the other for subcellular particles (CFP stained PML bodies). Strong deformations occur since the cells are going into mitosis (cell nucleus division). 9 spot-like structures from each 2D sequence and 6 from each 3D sequence in the nucleus channel were manually annotated and tracked.

The other two data sets “A” and “B” were acquired with a confocal microscope [10], [44], [45] (the data and annotations are available at <https://cbia.fi.muni.cz/CellRegistration>). In sequences A1 and A2, the nuclei of U2OS cells stained with

mCherry-BP1-2 were UV-irradiated in a stripe-like region. The nuclei undergo significant translation and rotation along with deformations. Key points of the UV-irradiated stripe were manually tracked by three independent annotators, considering that image noisy is rather significant and structures are blurry. Sequences B1-B4 depict HeLa cells with histone H2B tagged by GFP. In each sequence bleached regions form four intersecting line structures, which can be automatically detected (see Fig. 8). These line structures are stable w.r.t. the nucleus and represent the motion and deformation of the nucleus, and thus can be used for performance evaluation.

We also used synthetic image data for a direct evaluation of the estimated deformation fields. We generated four 2D synthetic sequences (G1-G4). The simulated time-lapse image sequences of cell nuclei should be as realistic as possible and the used deformations should be coherent and consistent with the cell shape. To achieve this, we used deformation fields estimated by an independent registration approach (pair-wise local method [14]) from real data (sequences S1-S4) as ground truth for the deformations. The synthetic image sequences were generated using the first frame of the real data and progressively deforming this frame by the ground truth deformation. The sequences comprise between 100 and 200 time points. Note that the ground truth deformation fields represent large deformations. To verify this we computed the mean and standard deviation (STD) of the magnitude of the deformation vectors between the first and the last frame for each synthetic sequence (see Tab. VII, second and third rows). It can be seen that the mean and STD are up to 44.3 pixels and 23.8 pixels, respectively, and the values differ largely for the different sequences, which corresponds to different levels of complexity of the deformations.

B. Implementation Details

For our global non-rigid registration method we employ a standard coarse-to-fine strategy for optimization to deal with large deformations. Gaussian pyramids with 5 scales and a downsampling factor of 0.5 are used. At each image scale, the deformation field is initialized using the result from the coarser scale, and then iteratively updated 5 times using the increment calculated between an image and the warped subsequent image (by the current estimated deformation field).

To train the two generative image models FoE and cGRBM and determine the features, we use 2000 image patches of size 50×50 pixels, which are randomly selected (excluding patches that do not contain any cell nuclei) from the nucleus channel in data set “S”. Since the images in data sets “A” and “B” contain more significant noise and rather blurry structures, which degrades the learning result, we exclude them from the training images. We have trained eight 2D filters of size 5×5 pixels for each model (see Fig. 3). It can be seen that, while the learned cGRBM filters appear better spatially structured, the FoE filters are more complex and more difficult to interpret. 3D filters were not used in our experiments due to limited data of data set “T” for learning 3D image models.

For the adaptive weighting and CLG schemes we use empirically determined fixed parameters $b = 10$ and $\sigma = 1$

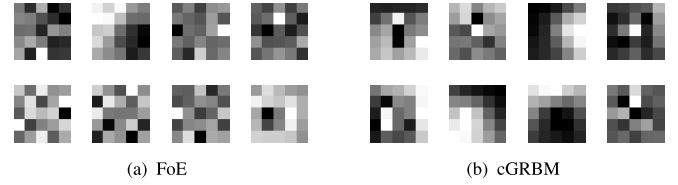


Fig. 3. Learned 2D filters (5×5 pixels) from two generative image models. (a) FoE. (b) cGRBM.

in (13), and $\rho = 1$ in (15) in all experiments for both real and synthetic images.

C. Performance Evaluation Using Real Image Data

1) *Geometric Registration Error*: Since ground truth of the deformation fields for the real microscopy image sequences is not available, it is not possible to directly evaluate the estimated deformation fields for registration. Evaluation of the intensity similarity between registered images is also inappropriate due to significant noise, intensity changes over time as well as appearing or vanishing structures. Instead, we evaluate the proposed methods by the geometric registration error as the Euclidean distance of an image structure in the registered image to its corresponding location in the reference image.¹

$$e_{mt} = \|\mathbf{p}_{m0} - \mathbf{p}_{mt}(\mathbf{w}^{0,t})\|_2, \quad (24)$$

where \mathbf{p}_{m0} denotes the coordinates of the m th image structure in the reference image and $\mathbf{p}_{mt}(\mathbf{w}^{0,t})$ represents those in the image at time point t transformed by the estimated deformation $\mathbf{w}^{0,t}$. We use manually or automatically annotated image structures in each sequence and compute the mean and standard deviation of the registration errors for these structures.

2) *Comparison of Different Regularizers*: We first compare basic global models (assuming brightness constancy, see Sec. II-B) with three different regularizers, namely quadratic, Charbonnier (with $\epsilon = 0.01$) and Lorentzian (with $\sigma = 0.1$). As the regularization parameter α_0 is the only free model parameter (see (5)) and its optimal value depends on the type of deformation field, we applied the models to all sequences in data set “S” with different values of α_0 . As can be seen in Fig. 4, the quadratic regularizer not only yields the lowest registration error, but is also more robust than the others due to its flatter error curve. This is consistent with our analysis in Sec. II-B and confirms that strong discontinuities in the deformation fields are unlikely.

3) *Comparison of Multiple Model Variants*: We also compared the performance of multiple variants of our global method (using a quadratic regularizer) for data set “S”: Models based on brightness constancy (“Br”) without and with adaptive weighting (“+AW”) or combined local-global (“+CLG”), model based on first-order derivative (gradient) constancy (“Deriv”), and models based on high-order feature constancy (“FoE” using learned FoE features, and “cGRBM” using

¹An exception is that, for line features in the sequences B1-B4, the Fréchet distance is computed as the registration error as in [10].

TABLE II
REGISTRATION ERRORS AND STANDARD DEVIATIONS (IN PIXEL) FOR ANNOTATED SPOT-LIKE STRUCTURES IN
2D IMAGE SEQUENCES S1-S4. RESULTS FOR PROPOSED GLOBAL MODELS AND PREVIOUS METHODS

Sequence	S1	S2	S3	S4	Avg.	Avg. STD
Unregistered	6.56	13.97	26.65	27.84	18.76	9.53
Local, multi-frame weighting method [13]	2.99	2.45	3.56	5.31	3.58	2.11
Local, diffeomorphic multi-frame method [14]	2.10	1.38	2.19	2.93	2.15	1.64
Contour-based method [10]	-	-	6.73	-	-	-
Our global model, image brightness “Br”	1.31	0.95	1.74	1.76	1.44	0.89
Our global model, image brightness “Br+AW”	1.36	1.03	1.86	1.87	1.53	0.97
Our global model, image brightness “Br+CLG”	1.28	0.95	1.76	1.58	1.39	0.87
Our global model, cGRBM features “cGRBM”	1.10	0.78	1.73	1.02	1.16	0.68
Our global model, cGRBM features “cGRBM+AW”	1.12	0.81	1.73	0.99	1.16	0.69
Our global model, cGRBM features “cGRBM+CLG”	1.05	0.79	1.83	0.89	1.14	0.67
Our global model, FoE features “FoE”	1.04	0.78	1.73	0.89	1.11	0.58
Our global model, FoE features “FoE+AW”	1.06	0.80	1.79	0.87	1.13	0.59
Our global model, FoE features “FoE+CLG”	0.94	0.78	1.78	0.73	1.06	0.66
Our global model, FoE features “FoE+AW+CLG”	0.92	0.78	1.76	0.74	1.05	0.66

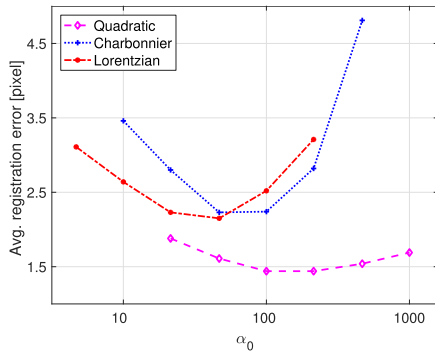


Fig. 4. Average registration error (in pixel) as a function of the regularization weight α_0 for global models based on brightness constancy and different regularizers for data set “S” (S1-S4).

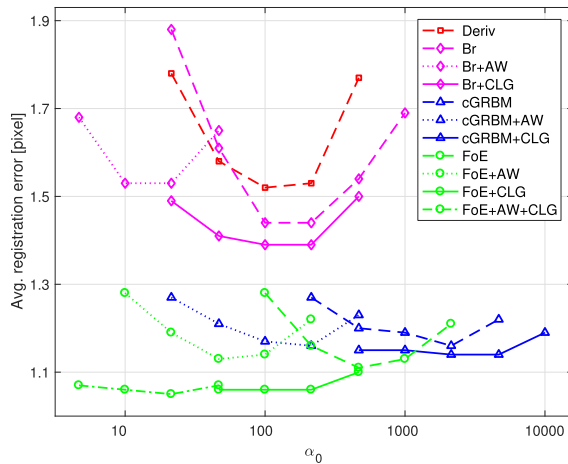


Fig. 5. Average registration error as a function of the regularization weight α_0 for different model variants of our global method for data set “S” (S1-S4).

learned cGRBM features) without and with adaptive weighting and/or CLG.

Fig. 5 shows the average registration errors of these models w.r.t. the regularization parameter α_0 . It can be seen that the model based on gradient constancy (“Deriv”) performs worse than that based on brightness constancy (“Br”), which suggests

that first-order derivative features are insufficient to capture structures in noisy microscopy images. Using the learned FoE or cGRBM features significantly improves the result and also increases the robustness w.r.t. α_0 , which can be attributed to the more accurate modeling of the data term using learned image features. For FoE we also tested the impact of the scheme for shifting the filter coefficients in our approach (*cf.* Sec. III-A.3 above). Without using the scheme, we obtained for S1-S4 an average registration error of 1.13 pixels with an average STD of 0.65 pixels, while using the scheme yielded a lower average error of 1.11 pixels and a lower average STD of 0.58 pixels. For this data set, the CLG scheme further improves the performance, while the AW scheme does not yield an improvement. This is not surprising since the noise and outliers in the images have a more complex pattern, but adaptive weighting assumes Poisson-Gaussian noise. The combined FoE + AW + CLG scheme yields a slightly lower average error compared to CLG. In Tab. II the registration errors and standard deviations of different variants of our global model with optimal α_0 (w.r.t. the whole data set) for all sequences of “S” are provided. The models using learned FoE features are slightly better than (or similar to) those using cGRBM features. An example of registration results for the sequence S4 can be seen in Fig. 6. For our global model “FoE + CLG” more yellow regions in the overlay of registered images are observable, which indicates a higher registration accuracy.

4) *Comparison With Existing Registration Methods:* We also performed a comparison with two local methods (local multi-frame weighting [13], local diffeomorphic multi-frame [14]) and a recent contour-based method [10]. For a fair comparison between our global models and previous methods, for data sets “S” and “T”, the model parameter α_0 is optimized for the whole data set (as done for the local models in [13] and [14]), and for data sets “A” and “B”, α_0 is optimized for every sequence (as for the contour-based method in [10]). Below, we present results using the optimal value of α_0 obtained by grid search.

For all sequences of data set “S”, it can be seen in Tab. II that all our global models outperform previous local methods [13], [14] and the contour-based approach [10].

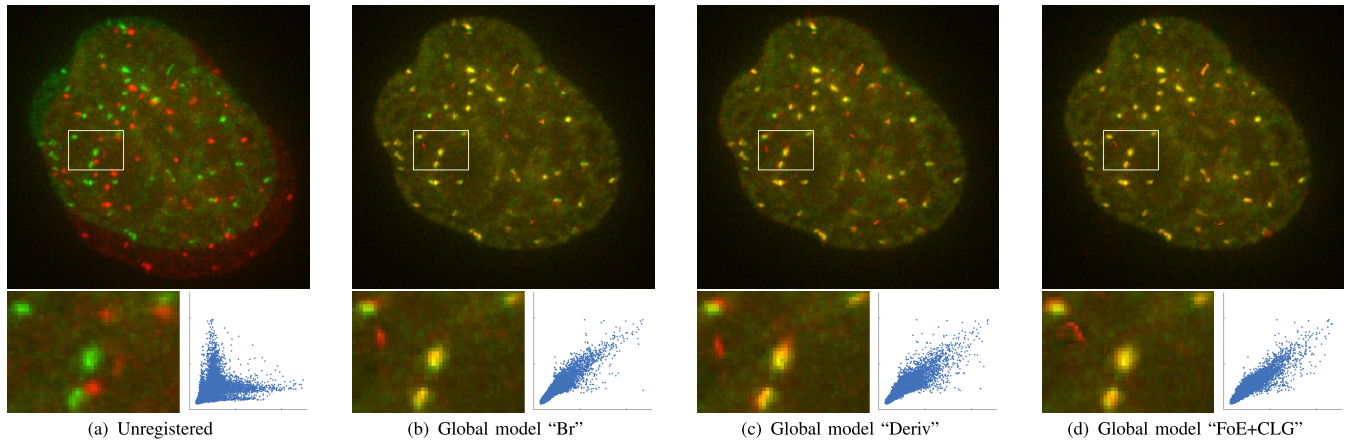


Fig. 6. Overlay and scatterplot of registered 20th frame (green) with 1st frame (red) of sequence S4, using different model variants of our global method. (a) Unregistered. (b) Global model "Br". (c) Global model "Deriv". (d) Global model "FoE + CLG".

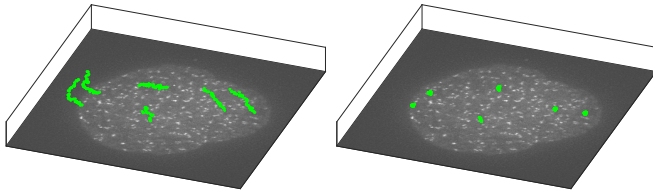


Fig. 7. Annotated spot-like structures (green dots) in 3D sequence T1. Unregistered (left) and registered (right) by our global model "Br + CLG".

TABLE III

REGISTRATION ERRORS AND STANDARD DEVIATIONS (IN PIXEL) FOR ANNOTATED STRUCTURES IN TWO 3D IMAGE SEQUENCES. RESULTS FOR PROPOSED GLOBAL MODELS AND PREVIOUS METHODS

Sequence	T1	T2	Avg.	Avg. STD
Unregistered	24.94	21.24	23.09	11.69
Multi-frame weighting [13]	6.98	4.63	5.81	3.75
Diffeomorphic multi-frame [14]	4.90	2.41	3.66	2.66
Our global model "Br"	1.83	1.43	1.63	0.96
Our global model "Br+CLG"	1.67	1.27	1.47	0.83

The best result is obtained by the model "FoE + AW + CLG" based on learned FoE features with a combined AW and CLG scheme for handling noise and outliers. The average registration error is decreased by more than 50% compared to [14].

Tab. III shows the results for the 3D image sequences of data set "T". As trained 3D image features are not available, only the results for our global model using brightness ("Br" and "Br + CLG") are provided. Compared to the local models in [13] and [14], our global model achieves a significant decrease of the registration error (75% and 60%, respectively). Fig. 7 shows the annotated spot-like structures for sequence T1 after registration by our global model "Br + CLG" compared to the unregistered case, which illustrates the registration accuracy (much smaller variation of the spot locations).

Sequences in data set "A" are more difficult to register due to significant noise, blurry image structures as well as

TABLE IV

REGISTRATION ERRORS (IN PIXEL) FOR ANNOTATED STRUCTURES IN SEQUENCES OF DATA SET "A"

Sequence	A1				A2			
	H_1	H_2	H_3	Avg.	H_1	H_2	H_3	Avg.
Unregistered	30.17	30.75	31.29	30.74	49.42	50.20	49.05	49.55
Cont.-based [10]	5.71	7.18	5.78	6.22	7.95	9.10	8.78	8.61
"Br+CLG"	6.87	7.60	5.63	6.70	8.47	8.90	9.60	8.99
"Br+AW"	6.69	7.42	5.38	6.50	7.19	7.85	7.80	7.61
"cGRBM+CLG"	7.27	7.19	5.37	6.61	7.77	8.18	7.88	7.94
"cGRBM+AW"	6.23	6.86	5.14	6.08	7.26	7.74	7.68	7.56
"FoE+CLG"	6.01	6.53	4.69	5.74	7.47	7.67	8.65	7.93
"FoE+AW"	5.98	6.81	3.93	5.57	6.45	6.78	7.86	7.03
"FoE+AW+CLG"	5.69	6.47	4.26	5.47	6.71	6.95	7.87	7.18

large intensity variation. The results are presented in Tab. IV. It can be seen that our global model generally yields better results than the previous contour-based approach [10]. The best result is obtained for the global model with learned FoE features, for which the average registration error is about 15% lower than that of [10]. It also turns out that the adaptive weighting scheme is superior to the CLG scheme for this data set. The reason is that adaptive weighting more appropriately models the image noise (significant noise but less outliers due to appearing or vanishing structures). Using FoE + AW + CLG improves the result for one image sequence, but the average error for both sequences is slightly worse compared to AW.

Next we consider data set "B". For performance evaluation we follow [10] and calculate the Fréchet distance between corresponding line features, and the Euclidean distances between corresponding intersection points and end points of the line structures. Fig. 8 shows an example. The quantitative results in Tab. V show that our global model yields better results than the previous contour-based approach [10]. Using learned features for FoE or cGRBM improves the result compared to the global model with brightness, while the cGRBM features perform slightly better than FoE features. Using adaptive weighting or CLG the result is not further improved. This is probably due to the bleached regions which significantly alter the image structure.

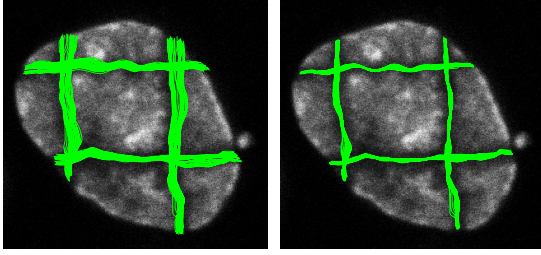


Fig. 8. Distributions of line structures in sequence B4. Unregistered (left) and registered (right) by our global model “cGRBM”.

TABLE V
REGISTRATION ERRORS (IN PIXEL) FOR ANNOTATED
STRUCTURES IN SEQUENCES OF DATA SET “B”

Sequence	B1	B2	B3	B4	Avg.
Line features					
Unregistered	22.91	11.54	12.43	11.92	14.70
Contour-based [10]	9.72	7.82	5.32	6.93	7.45
Our “Br”	10.05	8.20	4.75	6.46	7.36
Our “FoE”	8.82	7.87	5.79	6.51	7.25
Our “cGRBM”	8.97	7.82	4.94	6.37	7.02
Intersection points					
Unregistered	11.30	5.71	8.00	6.79	7.95
Contour-based [10]	7.92	5.22	2.76	3.04	4.73
Our “Br”	7.56	5.40	2.34	2.90	4.55
Our “FoE”	6.51	4.80	2.58	3.06	4.24
Our “cGRBM”	6.71	4.99	2.35	2.85	4.22
End points					
Unregistered	18.15	6.40	9.30	8.13	10.50
Contour-based [10]	3.69	2.02	2.17	2.43	2.58
Our “Br”	5.84	1.69	1.81	1.93	2.82
Our “FoE”	3.58	1.56	1.92	2.08	2.28
Our “cGRBM”	3.36	1.42	1.82	1.86	2.12

TABLE VI
COMPUTATION TIME (IN SECONDS) OF DIFFERENT MODEL VARIANTS
AND EXTRA COMPUTATION TIME COMPARED TO “Br” FOR
REGISTERING THE FIRST 100 IMAGES OF SEQUENCE S2

Br	Br+AW	Br+CLG	FoE	FoE+AW	FoE+CLG	FoE+AW+CLG
74	78 (+5%)	77 (+4%)	121 (+64%)	126 (+70%)	140 (+89%)	140 (+89%)

5) Computation Time: We have implemented our approach using Matlab without optimization and performed the experiments on a Linux workstation with an Intel Xeon E5 (2.90GHz) CPU. As the inference of our model involves solving a series of linear systems of equations with sparse coefficient matrix, computation is very efficient. Registration of sequence A2 (all 38 images) takes 1min37sec by our global method “FoE + AW”, while the contour-based approach [10] requires 3min46sec. Registration of the first 100 images of sequence S2 takes 1min14sec to 2min20sec depending on the model variant of our method (see Tab. VI for details). By contrast, the local method [14] needs 1h37min to 2h41min depending on the method variant. For the 3D data, while [14] requires more than 27h, our method only needs 8min.

TABLE VII
MEAN ENDPOINT ERRORS (IN PIXEL) OF THE ESTIMATED
DEFORMATIONS FOR FOUR SYNTHETIC IMAGE SEQUENCES
(WITHOUT NOISE), AND MEAN AS WELL AS STD
OF THE MAGNITUDE OF THE GROUND
TRUTH DEFORMATION VECTORS

Sequence	G1	G2	G3	G4	Avg.
Mean of vector magnitude	11.9	18.8	44.3	32.7	-
STD of vector magnitude	5.68	8.42	23.8	17.5	-
Multi-frame local method [14]	0.75	1.07	1.89	2.06	1.44
Our global model “Br”	0.47	0.49	0.95	1.25	0.79
Our global model “cGRBM”	0.44	0.48	0.77	1.20	0.72
Our global model “FoE”	0.42	0.47	0.80	1.11	0.70

D. Performance Evaluation Using Synthetic Data

Since for the synthetic data the ground truth deformations are known, we evaluate the proposed methods by the endpoint error (EE) defined as the Euclidean distance between an estimated deformation vector and the corresponding ground truth vector.

Tab. VII shows the mean EEs (over all pixels within the region of the cell nuclei and over all time points) of the estimated deformation fields for the four synthetic image sequences (G1-G4). Our global models yield a much better result than the multi-frame local model in [14]. Note that the sequences comprise different levels of complexity of the deformations (see Tab. VII, second and third rows for the mean and STD of the magnitude of the deformation vectors). Compared to [14], our model using FoE features decreases the mean EE by more than 50%. As the original synthetic images were generated without adding noise, we did not apply the model variants with AW or CLG.

To evaluate the robustness of our global models w.r.t. different spatial and temporal resolutions, we downsample the synthetic sequences spatially and temporally by a factor of 1/2. From Tab. VIII it can be seen that for 1/2 spatial downsampling, the mean EE is only slightly increased, while for 1/2 temporal downsampling, the mean EE is even decreased, since fewer image frames are used. In addition, the results for the different variants of the global model are consistent with those for the original synthetic data (see Tab. VII).

We also investigate the robustness of the proposed adaptive weighting (AW) scheme against image noise. We add mixed Poisson-Gaussian (PG) noise, which is typical for fluorescence microscopy images, to the synthetic image sequences with different noise levels according to $x' = st + n_N$ and $t \sim \text{Poi}(x/s)$, where x and x' denote intensity values of noise-free and noisy images, s is a gain factor controlling the Poisson noise strength, and n_N is additive Gaussian noise with variance σ_N^2 (example images are shown in Fig. 9). The registration results are provided in Tab. IX. It can be seen that the proposed adaptive weighting scheme consistently improves the registration accuracy (although we used empirically determined parameter values without knowing the exact noise levels, cf. Sec. IV-B), while the CLG scheme even slightly degrades the performance for strong noise. This observation is consistent with the results for the data set “A” of real images which include significant

TABLE VIII
MEAN ENDPOINT ERRORS (IN PIXEL) OF THE ESTIMATED
DEFORMATIONS FOR FOUR DOWNSAMPLED
SYNTHETIC SEQUENCES

Sequence downsampling	Original	1/2 spatial	1/2 temporal
Our model “Br”	0.79	0.96	0.69
Our model “cGRBM”	0.72	0.94	0.68
Our model “FoE”	0.70	0.91	0.65

TABLE IX
MEAN ENDPOINT ERRORS (IN PIXEL) OF THE ESTIMATED
DEFORMATIONS FOR FOUR SYNTHETIC SEQUENCES
WITH DIFFERENT LEVELS OF NOISE

Noise level	Gaussian $\sigma_N = 3$	Poisson $s = 0.3, \sigma_N = 0$	Mixed PG $s = 0.8, \sigma_N = 3$	Mixed PG $s = 1.2, \sigma_N = 5$
“Br”	1.58	2.02	3.16	3.87
“Br+CLG”	1.50	1.97	3.18	3.95
“Br+AW”	1.43	1.82	2.98	3.71
“Br+AW+CLG”	1.38	1.77	2.96	3.80

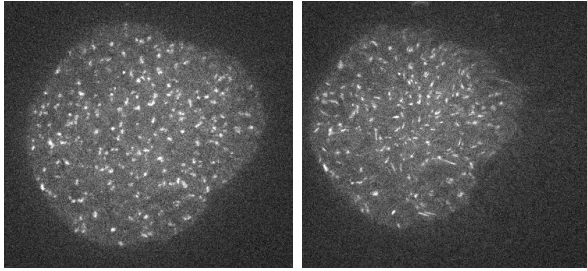


Fig. 9. Images from the synthetic sequence G3, with mixed Poisson-Gaussian noise ($s = 1.2, \sigma_N = 5$) for $t = 0$ (left) and $t = 50$ (right).

noise. The combined AW + CLG scheme improves the result compared to CLG and AW except for a high level of mixed PG noise, where AW is best.

E. Discussion

From experiments it turned out that our global models for non-rigid registration of cell microscopy images generally outperform existing methods. Best results are achieved by our global models using learned image features (FoE or cGRBM). In general, when the image structures are not artificially altered (e.g., by bleaching as for data set “B”), the registration error using FoE features is lower compared to cGRBM features. The learned FoE image model can better capture the statistics of training images, though FoE filters are less spatially structured and difficult to interpret (see Fig. 3). The adaptive weighting (AW) and CLG schemes generally further improve the registration accuracy. For images with significant noise (e.g., data set “A” and synthetic data with Poisson-Gaussian noise), AW is more suitable as the statistics of the noise is considered; for images containing many outliers (e.g., data set “S”, where irregular spot-like structures may vanish or arise), CLG yields better results as neighborhood information is utilized.

In practice, it is generally difficult to decide whether AW or CLG is better suited for certain data. If the noise statistics are known (e.g., significant noise or not), one could decide based on this information. Otherwise, a small amount of annotated data could be used to decide which of the two schemes should be applied. Alternatively, the combined

AW + CLG scheme can be used. We found that the combined scheme improves the result for certain image sequences, and the average error for the data sets is similar to CLG or AW, while the computation time is nearly the same as for CLG (see Tab. VI). An advantage of the combined scheme is that it is not necessary to decide between AW and CLG.

In our experiments we used different types of ground truth depending on the data set. For data set “S” manual annotation was used, which is relatively accurate since localizing spot-like structures and finding correspondences are relatively easy. Data set “A” contains significant noise and blurry structures, thus the annotation was performed by three independent annotators to increase the reliability. For data set “B”, intersecting line structures of bleached regions were automatically detected and used for the evaluation, since these line structures are found to be stable w.r.t. the nucleus. However, the artificial line structures due to bleaching change the properties and statistics of cell microscopy images. As the bleached regions are dark and lack spatial structures, local deformations are hardly captured within these regions. Therefore, the results for data set “B” should be considered with caution.

In our global model we have employed an MRF as the regularizer of the deformation fields, which is not only effective, but also allows arbitrary deformations. Although the regularizer can be straightforwardly extended by additional terms, one should be careful using other constraints or physical prior (e.g., elasticity), because they might be too strict in the case of vanishing or appearing structures in real data.

FoE and cGRBM used for learning features in our global model have a sound probabilistic interpretation and the learned filters can capture important features of the considered images. Their convolutional model structure is not only consistent with the (hidden) independence assumption in the registration model, but also requires moderate amount of training data, which is important for many biological problems with limited data. In our experiments, we found that the models using FoE features yield slightly better registration results than those using cGRBM features, however, training of FoE features is slower compared to cGRBM features. Integration of interpretable high-level features learned, for example, by deep models, into our global registration model is interesting to explore, but this would require significant extension of the proposed framework. This will be a subject of future work.

V. SUMMARY

We introduced global optical flow models for non-rigid registration of time-lapse fluorescence microscopy images of cells. We investigated different regularizers and found that a convex quadratic function is best suited for regularization of the deformation fields. We also employed two generative image models, namely Fields of Experts and convolutional Gaussian restricted Boltzmann machine, to train high-order image features using an augmented learning procedure. The learned features were integrated into our global model to increase the registration accuracy. To improve the robustness to noise and outliers, we proposed an adaptive weighting scheme and a combined local-global scheme as well as a combination of these two schemes. Experiments using multiple fluorescence

microscopy data sets comprising 2D and 3D image sequences demonstrated that our proposed global method outperforms previous approaches. We also provided insights about the problem formulation and the used computational methods.

REFERENCES

- [1] B. Rieger, C. Molenaar, R. Dirks, and L. V. Vliet, "Alignment of the cell nucleus from labeled proteins only for 4D in vivo imaging," *Microsc. Res. Technol.*, vol. 64, no. 2, pp. 142–150, 2004.
- [2] A. P. Goobic, J. Tang, and S. T. Acton, "Image stabilization and registration for tracking cells in the microvasculature," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 2, pp. 287–299, Feb. 2005.
- [3] C. A. Wilson and J. A. Theriot, "A correlation-based approach to calculate rotation and translation of moving cells," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 1939–1951, Jul. 2006.
- [4] P. Matula, P. Matula, M. Kozubek, and V. Dvorak, "Fast point-based 3-D alignment of live cells," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2388–2396, Aug. 2006.
- [5] O. Dzyubachyk *et al.*, "Automated analysis of time-lapse fluorescence microscopy images: From live cell images to intracellular foci," *Bioinformatics*, vol. 26, no. 19, pp. 2424–2430, 2010.
- [6] S. Ozeré, P. Bouthemy, F. Spindler, P. Paul-Gilloteaux, and C. Kervrann, "Robust parametric stabilization of moving cells with intensity correction in light microscopy image sequences," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2013, pp. 468–471.
- [7] J. Mattes, J. Nawroth, P. Boukamp, R. Eils, and K. M. Greulich-Bode, "Analyzing motion and deformation of the cell nucleus for studying colocalizations of nuclear structures," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2006, pp. 1044–1047.
- [8] S. Yang, K. Teller, D. Koehler, and T. Cremer, "Nonrigid registration of 3-D multichannel microscopy images of cell nuclei," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 493–499, 2008.
- [9] J. De Vylder, W. H. De Vos, E. M. Manders, and W. Philips, "2D mapping of strongly deformable cell nuclei-based on contour matching," *Cytometry A*, vol. 79A, no. 7, pp. 580–588, 2011.
- [10] D. V. Sorokin, I. Peterlik, M. Tektonidis, K. Rohr, and P. Matula, "Non-rigid contour-based registration of cell nuclei in 2-D live cell microscopy images using a dynamic elasticity model," *IEEE Trans. Med. Imag.*, vol. 37, no. 1, pp. 173–184, Jan. 2018.
- [11] I. Kim, Y. M. Chen, D. L. Spector, R. Eils, and K. Rohr, "Non-rigid registration of 2-D and 3-D dynamic cell nuclei images for improved classification of subcellular particle motion," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1011–1022, 2011.
- [12] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. J. Conf. Art. Intel. (IJCAI)*, Apr. 1981, pp. 674–679.
- [13] M. Tektonidis, I.-H. Kim, Y.-C. M. Chen, R. Eils, D. L. Spector, and K. Rohr, "Non-rigid multi-frame registration of cell nuclei in live cell fluorescence microscopy image data," *Med. Image Anal.*, vol. 19, no. 1, pp. 1–14, 2015.
- [14] M. Tektonidis and K. Rohr, "Diffeomorphic multi-frame non-rigid registration of cell nuclei in 2d and 3d live cell images," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1405–1417, 2017.
- [15] D. Fortun, P. Bouthemy, and C. Kervrann, "Optical flow modeling and computation: A survey," *Comput. Vis. Image Understand.*, vol. 134, pp. 1–21, May 2015.
- [16] D. Sun, S. Roth, J. P. Lewis, and M. J. Black, "Learning optical flow," in *Proc. Eur. Conf. Comp. Vis. (ECCV)*, 2008, pp. 83–97.
- [17] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [18] M. Welling, G. E. Hinton, and S. Osindero, "Learning sparse topographic representations with products of student-t distributions," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2003, pp. 1383–1390.
- [19] C. Metz, S. Klein, M. Schaap, T. V. Walsum, and W. J. Niessen, "Nonrigid registration of dynamic medical imaging data using nD+tB-splines and a groupwise optimization approach," *Med. Image Anal.*, vol. 15, no. 2, pp. 238–249, 2011.
- [20] M. Yigitsoy, C. Wachinger, and N. Navab, "Temporal groupwise registration for motion modeling," in *Proc. Int. Conf. Inf. Process. Med. Imag. (IPMI)*, 2011, pp. 648–659.
- [21] S. Durrleman, X. Pennec, A. Trounev, J. Braga, G. Gerig, and N. Ayache, "Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data," *Int. J. Comput. Vis.*, vol. 103, no. 1, pp. 22–59, 2013.
- [22] W. Shi *et al.*, "Temporal sparse free-form deformations," *Med. Image Anal.*, vol. 17, no. 7, pp. 779–789, 2013.
- [23] M. Polfliet, S. Klein, W. Huizinga, M. M. Paulides, W. J. Niessen, and J. Vandemeulebroucke, "Intrasubject multimodal groupwise registration with the conditional template entropy," *Med. Image Anal.*, vol. 46, pp. 15–25, May 2018.
- [24] U. Schmidt, Q. Gao, and S. Roth, "A generative perspective on MRFs in low-level vision," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1751–1758.
- [25] Q. Gao and S. Roth, "How well do filter-based MRFs model natural images?" in *Proc. DAGM-Symp. Pattern Recognit.* in Lecture Notes in Computer Science, vol. 7476. Berlin, Germany: Springer, 2012, pp. 62–72.
- [26] H. Zhang, Y. Zhang, H. Li, and T. S. Huang, "Generative bayesian image super resolution with natural image prior," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4054–4067, Sep. 2012.
- [27] K. Schelten, S. Nowozin, J. Jancsary, C. Rother, and S. Roth, "Inter-leaved regression tree field cascades for blind image deconvolution," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Jan. 2015, pp. 494–501.
- [28] U. Schmidt, J. Jancsary, S. Nowozin, S. Roth, and C. Rother, "Cascades of regression tree fields for image restoration," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 38, no. 4, pp. 677–689, Apr. 2016.
- [29] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," Univ. Toronto, Dept. Comput. Sci., Ottawa, ON, Canada, Tech. Rep. UTML TR 2010-003, 2010.
- [30] Q. Gao and S. Roth, "Texture synthesis: From convolutional RBMs to efficient deterministic algorithms," in *Structural, Syntactic, and Statistical Pattern Recognition* (Lecture Notes in Computer Science), vol. 8621. Berlin, Germany: Springer, 2014, pp. 434–443.
- [31] G. V. Tulder and M. De Bruijne, "Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted boltzmann machines," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1262–1272, May 2016.
- [32] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/kanade meets horn/schunck: Combining local and global optic flow methods," *Int. J. Comput. Vis.*, vol. 61, no. 3, pp. 211–231, Feb. 2005.
- [33] Q. Gao and K. Rohr, "Optical flow-based non-rigid registration of cell nuclei: Global model with adaptively weighted regularization," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 420–423.
- [34] Y. Tseng, J. S. H. Lee, T. P. Kole, I. Jiang, and D. Wirtz, "Micro-organization and visco-elasticity of the interphase nucleus revealed by particle nanotracking," *J. Cell Sci.*, vol. 117, pp. 2159–2167, Jan. 2004.
- [35] M. J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Comput. Vis. Image Understand.*, vol. 63, no. 1, pp. 75–104, Jan. 1996.
- [36] Y. W. Teh, M. Welling, S. Osindero, and G. E. Hinton, "Energy-based models for sparse overcomplete representations," *J. Mach. Learn. Res.*, vol. 4, pp. 1235–1260, Dec. 2003.
- [37] S. Roth and M. J. Black, "Fields of experts," *Int. J. Comput. Vis.*, vol. 82, no. 2, pp. 205–229, Apr. 2009.
- [38] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *Proc. IEEE Comput. Soc. Conf. Comp. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2432–2439.
- [39] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J.-B. Sibarita, and J. Salamero, "Patch-based nonlocal functional for denoising fluorescence microscopy image sequences," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 442–454, Feb. 2010.
- [40] J. O. Irwin, "The frequency distribution of the difference between two independent variates following the same poisson distribution," *J. Roy. Statist. Soc.*, vol. 100, no. 3, pp. 415–416, 1937.
- [41] R. Barrett *et al.*, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. Philadelphia, PA, USA: SIAM, 1994.
- [42] M. Norouzi, M. Ranjbar, and G. Mori, "Stacks of convolutional restricted boltzmann machines for shift-invariant feature learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 2735–2742.
- [43] H. Lee, R. Grosse, R. Ranganath, and A. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Jun. 2009, pp. 609–616.
- [44] D. V. Sorokin, J. Suchánková, E. Bártošová, and P. Matula, "Visualizing stable features in live cell nucleus for evaluation of the cell global motion compensation," *Folia Biologica*, vol. 60, pp. 45–49, Jan. 2014.
- [45] V. Foltánková, P. Matula, D. Sorokin, S. Kozubek, E. Bártošová, "Hybrid detectors improved time-lapse confocal microscopy of PML and 53BP1 nuclear body colocalization in DNA lesions," *Microsc. Microanal.*, vol. 19, no. 2, pp. 360–369, 2013.