

Automatic Tracking of Escherichia Coli Bacteria

Jun Xie¹, Shahid Khan^{2,3}, and Mubarak Shah⁴

¹ Janelia Farm Research Campus, HHMI, USA
xiej@janelia.hhmi.org

² Molecular Biology Consortium, Chicago, USA

³ LUMS_SSE, Lahore, Pakistan

⁴ Computer Vision Lab, University of Central Florida, USA

Abstract. In this paper, we present an automatic method for estimating the trajectories of *Escherichia coli* bacteria from *in vivo* phase-contrast microscopy videos. To address the low-contrast boundaries in cellular images, an adaptive kernel-based technique is applied to detect cells in sequence of frames. Then a novel matching gain measure is introduced to cope with the challenges such as dramatic changes of cells' appearance and serious overlapping and occlusion. For multiple cell tracking, an optimal matching strategy is proposed to improve the handling of cell collision and broken trajectories. The results of successful tracking of *Escherichia coli* from various phase-contrast sequences are reported and compared with manually-determined trajectories, as well as those obtained from existing tracking methods. The stability of the algorithm with different parameter values is also analyzed and discussed.

1 Introduction

The study of cell movement in response to chemical and environmental agents has been an important research area in the bio-medical and environmental science community for quite some time [1,2]. Biologists typically need manual or interactive computer-assisted tracking of cell motion to study chemotactic responses. Manual tracking becomes impractical for data sets where thousands of cells are involved. Automated tracking and analysis of the cells' motility thus becomes critical for time-resolved analysis of the underlying biological mechanisms.

The objective of this study is to track from microscopy videos the gram-negative organism *Escherichia coli* bacteria (*E. coli*), which can generally cause several intestinal and extra-intestinal infections such as urinary tract infections, meningitis, and peritonitis. *Escherichia coli* chemotaxis has been the system of choice for elucidation of the design principles of transmembrane and intracellular signal transduction. Automated tracking and motion analysis would significantly enhance the investigators' ability to study *E. coli*, improve data processing efficiency and remove operator bias.

Tracking typically consists of identifying unique objects in a complex environment where the background remains fairly constant and the target maintains a similar appearance. While numerous methods have been proposed for general object tracking [3], cellular videos pose many challenges to those existing techniques due to severe image noise and clutter, shape deformation, and high processing demand.

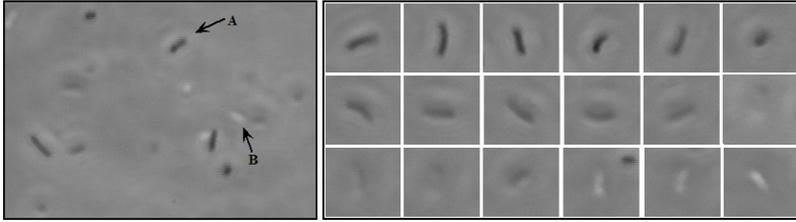


Fig. 1. (Left) A typical view of *E. coli* bacteria under a phase-contrast microscopy. The cell soma generally appears as a dark area surrounded with a white halo (A) when it is in the focal plane. Once it moves sufficiently out of focus, the contrast will be inverted and the same cell may appear as a white bulb (B). (Right) An sequence of *E. coli* bacteria.

Among the efforts devoted to cellular imaging, there is one class of methods that perform tracking using edge information [4,5]. Unfortunately, the close proximity of cells and occlusion in the *in vivo* microscopy videos make edge-based cell tracking difficult. A large number of adaptations are required for these methods to be successfully applied. Rather than segmenting the object precisely, some methods consider tracking as a problem of centroid relocation [1] to simplify the tracking task and avoid the requirement of boundary detection.

In order to track living *E. coli* from phase-contrast microscopy videos, we follow the idea of centroid tracking. Although there is no cell division in our case, serious collision and large overlapping pose more challenges to precise border detection. As shown in Fig. 1, the *E. coli* cells typically have a large range of motion patterns and the cell soma generally appears as a dark area surrounded with a white halo, but the contrast can be inverted and the cell appears white when it has moved sufficiently out of focus. Those facts make tracking based on the constancy in shape and intensity difficult. Padfield *et. al* [6] recently proposed to generate a dynamic model to describe the appearance change of nuclei over time for live cell tracking.

Another challenging issue in this specific tracking task is incomplete trajectories. Since individual *E. coli* bacteria can swim freely in 3D space, they may stray from the narrow focal plane and hence become temporarily lost, causing fragmentation of their trajectories. That is why we consider multi-cell tracking as a global optimal assignment problem.

2 Method

The proposed method starts with an object detection step which identifies the moving objects against the relative constant background. Then a global matching strategy is applied to estimate the cells' trajectories based on image appearance and motion patterns.

2.1 Cell Detection in Fuzzy Scene

In order to classify foreground and background, we apply our previous method [7] which is able to handle multiple objects with fuzzy edges. The method is based on

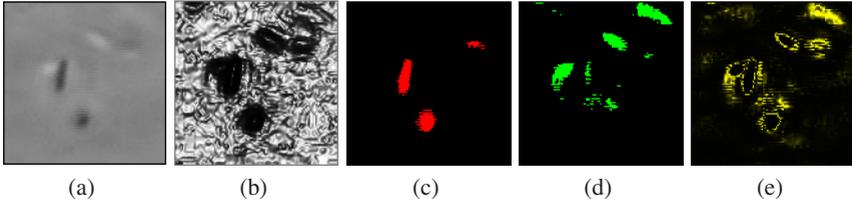


Fig. 2. An example of coli detection. (a) A patch from the original frame. (b) The homogeneity map computed using GVF measurement. (c) and (d) show the detected objects (without morphology operations). (e) The final entropy map.

the observation that the pixels with high class uncertainty accumulate mostly around object boundaries. We model the image as a mixture of Gaussians, and the optimal pixel classification is obtained by minimizing the loss function $L = L_{\text{entropy}} - \lambda L_{\text{likelihood}}$, where λ is a scale factor. $L_{\text{likelihood}} = \int_{\mathbf{x} \in \Omega_i} \log P(\mathbf{x})$ represents the likelihood and L_{entropy} is the entropy term defined as:

$$L_{\text{entropy}} = \int_{\mathbf{x} \in \Omega_i} (1 - G(\mathbf{x}))U(\mathbf{x}) + (1 - U(\mathbf{x}))G(\mathbf{x}), \quad (1)$$

where G is the normalized gradient vector flow (GVF) [8] serving as a measure of spatial information, and U is the normalized entropy describing the class uncertainty according to Shannon's theory. By minimizing the cost function, we can simultaneously optimize the parameters of the global model and the distribution of entropy for the detection process.

The optimization of those parameters can be achieved through the Quasi-Newton algorithm [9]. To improve the efficiency, we applied the EM method [10] at the initial stage to find the initial model parameters and the size of each object category. Figure 2 shows an example of the detection results. More details about this detection method can be found in [7]. After pixels are classified into different groups, the Connected Components Labeling technique [11] can be applied to generate connected regions (bulbs), and regions with sizes comparable with pre-selected thresholds are regarded as candidates.

2.2 Matching Gain for Candidate Selection

After detecting the candidates, an intuitive way to track the target, as applied in the Mean-Shift approach [12], is to compare the intensity similarity between the candidates and the targeted cells. Usually the intensity histogram will be employed to describe the intensity profile of each object.

In our context, where the appearance of the target changes very quickly (see Fig. 1 (Right) for an example), the single intensity similarity is not reliable enough to provide a robust measurement for the tracking. In addition, there are numerous cells moving in the field of view and interacting with each other so closely that it is difficult, even for human, to identify the correct tracks for those cells from one frame to the next. This is why we consider the global trajectory inference and promote the use of a graph based approach for optimal position estimation.

Multiple Cell Tracking: First, let's consider the simple case of multiple cells observed in two successive frames. Let c_i^p ($i = 1 \dots l$) denote detected cells in the p -th frame and c_j^{p+1} ($j = 1 \dots k$) represent those detected in the $(p + 1)$ -th frame. The task is to find the matching from $\{c_i^p\}$ to $\{c_j^{p+1}\}$. This context can be modeled with a complete bi-partite graph $G = (U, V, E)$ where $U = \{c_1^p, c_2^p \dots c_l^p\}$, $V = \{c_1^{p+1}, c_2^{p+1} \dots c_k^{p+1}\}$, and E represents the set of matching hypotheses between each pair of cells from frame p to $p + 1$. The correspondence problem can then be posed to find a matching of graph G , which is defined as a set of edges with no shared end-vertices. Assume there are pre-defined functions \mathcal{W} associated with each of these edges. A minimum matching in a weighted graph is a matching with minimum weight among all matchings in the graph. Since any two detected cells may hypothetically match, the resulting bi-partite graph is complete. Given the weights defined by specific matching criteria, a unique matching \widetilde{M} between the cells in two frames can be found as $\widetilde{M} = \arg \min_{M \in C} \sum \mathcal{W}(c_i^p, c_j^q)$, where C is the union of all the possible matching of G . There are several efficient algorithms (e.g., [13]) which can be used to find the minimum matching of a bipartite graph.

For multiple cells in multiple frames, it is a complete k -partite graph and this matching problem is NP-hard [14]. Fortunately, as demonstrated in [15], if the graph is an acyclic directed graph, a polynomial-time solution exists where the edges of the minimum matching of the split graph Ξ of an acyclic edge-weighted directed graph G correspond to the edges of minimum path cover of G . To model the multiple frame matching problem, we construct a weighted directed graph $G = (\{V_1, V_2, \dots, V_k\}, E)$, where V_i represents the set of detected cells in frame F_i . Each edge $e = (c_i^p, c_j^q)$ corresponds to a match hypothesis of coli c_i^p in frame F_i to coli c_j^q in frame F_j , and the weight w_e is defined as the matching cost, like in 2-frame cases.

If coli c_i^p has no correspondences in several consecutive frames and gets its forward matching $c_j^{p+\delta}$ in frame $F_{p+\delta}$ (due to occlusion or being out of focal plane), an edge $e = (c_i^p, c_j^{p+\delta})$ can naturally handle broken trajectories and thus provide an overall coverage of the possible solutions. Since all the edges in graph G are in the temporal direction, it is guaranteed that G is acyclic. The only requirement for this graph is that the weight function must satisfy the inequality $\mathcal{W}(c_i^p, c_j^q) < \mathcal{W}(c_i^p, c_{i+1}^{p+1}) \mathcal{W}(c_{j-1}^{q-1}, c_j^q)$ in order to penalize the choice of shorter trajectories when longer valid ones are present.

Matching Criteria: According to the discussion above, it is obvious that the weight function is critical for the correct matching of different cells across the long sequence. The weight of matching two cells c_i^p and c_j^q can be defined based on the cell appearances as follows:

$$\mathcal{W}_h(c_i^p, c_j^q) = \frac{1}{2} \sum_{b=1}^B \frac{[\mathbf{h}_i^p(b) - \mathbf{h}_j^q(b)]^2}{\mathbf{h}_i^p(b) + \mathbf{h}_j^q(b)}, \quad (2)$$

where \mathbf{h}_i^t refers to the intensity histogram of the i -th cell in frame t and B indicates the total gray levels.

An alternative way is to define the gain function based on some assumption of the undergoing motion of the targeted cell, such as a constant direction or velocity. The prediction of the cell in a new frame is then estimated accordingly. Then the matching

weight can be defined using motion measurements. A simple function for this is the nearest neighborhood criteria.

A matching function considering both direction and motion coherence has also been used in [16] for tracking feature points. The function is defined as follows:

$$\mathcal{W}(c_i^p, c_j^q) = \gamma \mathcal{W}_o + (1 - \gamma) \mathcal{W}_v, \quad (3)$$

where $\mathcal{W}_o = \left(\frac{1}{2} + \frac{\overrightarrow{c_i^p c_j^q} \cdot \overrightarrow{c_i^p c_j^q}}{2\|c_i^p c_j^q\| \|c_i^p c_j^q\|}\right)$ is the orientation term, and $\mathcal{W}_v = \frac{2\sqrt{\|c_i^p c_j^q\| \|c_i^p c_j^q\|}}{\|c_i^p c_j^q\| + \|c_i^p c_j^q\|}$ represents the velocity term which prefers the match with less change in the magnitude of velocity.

Although those motion-based measures enforce the coherence of motion, they require initialization of correspondence obtained manually or by other criteria. Also, the motion coherence may not be sufficient in the presence of highly random motions as in our application. In addition, they are not suited to handle cells entering the field of view late.

Based on the observation that the significant changes of the coli's appearance are usually accompanied by the occurrence of contrast corruption, we define our matching gain function as following:

$$\mathcal{W}(c_i^p, c_j^q) = (1 - \omega_i^p) \mathcal{W}_h(c_i^p, c_j^q) + \omega_i^p (\mathcal{W}_d(c_i^p, c_j^q) + \mathcal{W}_o(c_i^p, c_j^q)), \quad (4)$$

where $\mathcal{W}_d(c_i^p, c_j^q) = \left(1 - \frac{\|c_i^p c_j^q - c_i^p c_j^q\|}{\sqrt{w_l^2 + h_l^2}}\right)$ refers to the distance between the predicted and observed cells (w_l and h_l are the width and height of the frame, respectively). The term ω_i^p represents a scalar to balance the impact between the intensity and distance coherence. It is defined based on the contrast measurement of the targeted cell as follows:

$$\omega_i^p = \frac{\alpha (\bar{I}_i^p - \bar{I}_{\Omega_i})^2}{2 (\bar{I}_i^p + \bar{I}_{\Omega_i})}, \quad (5)$$

where α is a constant $0 \leq \alpha \leq 1$, \bar{I}_i^p and \bar{I}_{Ω_i} refers to the average image intensity of region covered by c_i^p and its local window Ω_i , respectively.

The above matching gain function is a convex combination of the intensity measurement (\mathcal{W}_h) and motion gains (\mathcal{W}_d and \mathcal{W}_o). They are adaptively combined based on the contrast of the tracked cell so that the intensity term will dominate the matching measurement when the target is clearly presented, while the motion clues will take over when the cell becomes blurry.

3 Experimental Results

In this section, we assess the proposed approach by comparing it with both popular tracking methods and manual tracking. Cultures for behavioral experiments were harvested at mid-exponential phase by centrifugation, washed three times, and re-suspended in a

Table 1. The Detection Validation and Tracking Performance on Each Sequence

Seq. No.	Res. (pixels)	Frames	Prec. (%)	Rec. (%)	MS	MSCE	MSAB	NEW
1	616 × 459	450	97.4	97.6	37.5	23.7	29.6	14.9
2	720 × 480	458	98.5	96.6	27.8	21.7	43.6	10.6
3	584 × 416	450	98.9	95.8	21.8	19.9	28.3	12.1
4	444 × 362	160	86.5	88.4	20.1	18.7	31.0	6.6
5	620 × 469	680	97.5	96.4	26.5	24.2	25.3	8.5
6	720 × 480	498	98.1	96.7	28.2	27.5	29.1	11.0
7	720 × 480	702	99.3	97.9	18.9	18.0	17.8	3.8
8	532 × 382	460	99.3	98.0	19.9	19.4	18.8	3.2
Avg.	620 × 441	482.2	96.94	95.93	25.09	21.64	27.93	8.83

potassium phosphate-EDTA motility buffer containing 5 mM lactate, as respiratory substrate, and 100 μ M methionine to maintain vigorous swim-tumble bias. The sequences were imaged by a CCD camera mounted on a Nikon Optiphot microscope using a phase contrast objective (40x CF Fluor plan-apochromat, 0.85 numerical aperture) and zoom lens [17]. Due to variations in shutter time, light exposure, filtering and cell culture, the data sets vary among themselves in contrast, intensity and apparent proximity. Our trial data consists of eight sequences (Table 1), totaling nearly 3800 frames that contain numerous *E. coli* cells moving naturally in the three-dimensional space.

One advantage of our approach is that it does not require manual initialization. Starting from the first frame, the regions of interest are detected in each frame as discussed in Section 2.1. To evaluate the detection accuracy, as applied in [5], we compute the *precision* as the ratio of the number of detected cells to the total number of detected candidates, and the *recall* as the ratio of detected cells to the total number of cells actually in the frame. In order to obtain the ground-truth for validation, a tool has been developed for operators to identify the cell centroid in each of the frames. The computed precision and recall for each individual sequence are listed in Table 1.

After detection, the intensity profile for each candidate is computed and its contrast measurement is estimated within a local window, which is double the size of the candidate. To establish the initial correspondence, the first two frames are used to construct a bipartite graph, where the weights are computed using Eq. (4) with $\alpha = 1$. The following frames are then processed sequentially using a constant α which is selected empirically ($\alpha = 0.2$ in our experiment).

To improve the efficiency of the algorithm, we compute the correspondence within a spatial window. Given the initial correspondence, we extend the graph by computing the weights for the successive frame. The minimum path cover of the graph is then estimated, and this procedure is repeated until a specific number of frames have been included. The size of the sliding window k affects the computation complexity and the capability of the algorithm to handle occlusion and broken trajectory. In order to correct the mismatches in previous procedure, a backtracking can be performed by applying the same tracking method in the reverse time direction as applied in [18,19]. Figure 3 shows the cell tracking results on two trial sequences.

We compared the tracking performance of the proposed algorithm with that achieved by human operator. Also, several Mean Shift (MS) technique based algorithms are

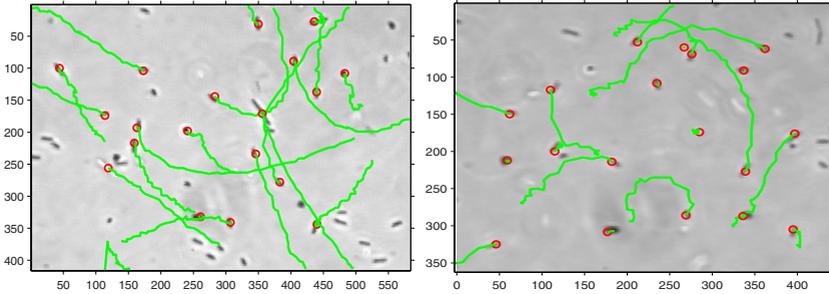


Fig. 3. Samples of estimated traces (red circles indicate start points)

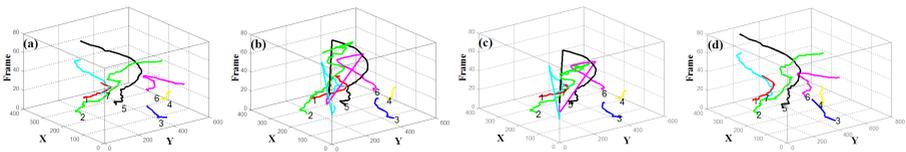


Fig. 4. Visualization of cell trajectories (in Sequence A) obtained using different methods: (a) Manual tracking, (b) Mean Shift, (c) MSCE, and (d) Our method. Each color curve shows a cell trajectory.

tested, including the classical Mean Shift [12], MS based on Contrast Enhancement (MSCE) [20], and MS using Adaptive Bandwidth (MSAB) [21]. For the proposed method, we assume there are three kernels for cell detection and the scale parameter λ was chosen as $1/N$ (N : the number of pixels in each frame). We also set $\alpha = 0.2$ and $k = 5$ for the test. For the classical Mean Shift algorithm, the histogram was generated using 64 levels and the model of the tracked cell was updated every 3 frames with a regression level of 0.3. Following the work in [20], we use the analysis resolution $\delta = 0.02$, outlier threshold $\tau = 0.01$ and distortion limit $\lambda = 5$ for the MSCE algorithm. As used in [21], we set the logarithmic coordinate base $b = 1$ and the scale level $s = 2$ in the MSAB method. The spatial-temporal plots in Fig. 4 demonstrate the three-dimensional views of several cells’ traces obtained with different methods. As noted, the Mean-Shift based techniques, compared to the manual results (Fig. 4(a)), generated some incorrect jumps (line 5(black), 6(pink) and 7(cyan)) due to poor contrast of the frames and the interference from cells with similar intensity. In contrast, the proposed approach (Fig. 4(d)) solved those problems smoothly and provided a solid performance, which is comparable to manual results.

In order to evaluate the tracking performance quantitatively, we apply the following criteria to measure the accuracy of the tracking algorithms. For each automatic cell trajectory, we computed the average distance (in pixels) in each frame between the manually-marked locations and those computed by the algorithms. If the distance in one frame is smaller than a preselected threshold (for example, the half of the cell’s size), the tracking result in this frame is considered to be correct and the frame is counted as a correct frame for the algorithm. Then the measure, called *frame-based error* [15], is

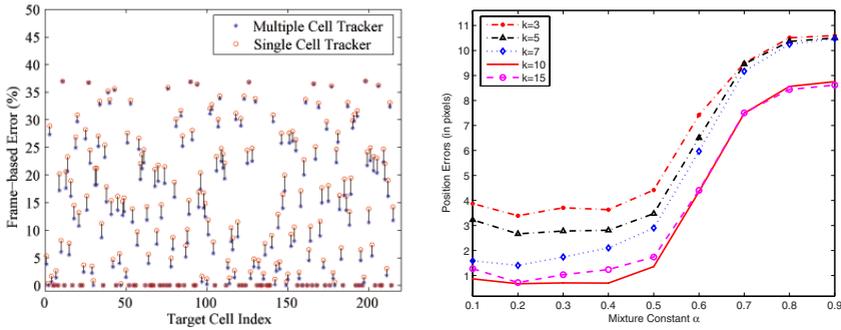


Fig. 5. (Left) The frame-based errors using the single-cell tracking method and multi-cell tracker. The horizontal axis represents the targeted cells. (Right) Tracking performance with respect to the mixture constant α : position errors (in pixels) for different size of sliding window k .

computed for the validation which is defined as $E_f = 1 - n_p/n_t$, where n_p is the number of correct frames generated by the tracker and n_t is the total number of frames the cell appears in the sequence.

The tracking errors of different tracking methods are illustrated in Table 1. In all the tested sequences, the proposed method has the best performance, with an average frame-based error of $8.83 \pm 4.10\%$. Figure 5(Left) shows the difference between the tracking results using the single cell tracker and multi-cell tracker, which confirms that the optimal matching strategy is able to improve the overall tracking accuracy. In Fig. 5(Right) we demonstrate the average pixel-based tracking errors with different values of the mixture constant α and the size of temporal window. Notice that typically the value of constant α will affect the tracking accuracy, and the better results are achieved with low values. The plot also shows that the tracking errors can be generally reduced by increasing the size of the sliding window. However, since a large window will increase the computation complexity, there is a trade-off between the tracking accuracy and computation time. For our case, the tracking system typically takes 3.3s/frame with a Matlab implementation on a 2GHz PC.

4 Conclusion

This paper introduced a fully automated method for tracking Escherichia coli bacteria in phase-contrast microscopy videos. The proposed detection method has been successfully applied to detect cells in the low contrast frame images. To handle the ambiguity in cellular images, a global optimal matching strategy is also introduced to enable multi-cell tracking. We have demonstrated the utility of the proposed algorithm for tracking E. coli bacteria from classical phase-contrast microscopy videos. Coupled with additional parameters for measurement of morphology, we anticipate that this algorithm will find wide application in diagnosis of bacterial pathogens in clinics and in basic biomedical research on bacterial chemotaxis.

References

1. Debeir, O., VanHam, P., Kiss, R., Decaestecker, C.: Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes. *IEEE Trans. on Medical Imaging* 24(6), 697–711 (2005)
2. Al-Kofahi, O., Radke, R., Goderie, S., Shen, Q., Temple, S., Roysam, B.: Automated cell lineage construction: a rapid method to analyze clonal development establish. *Cell Cycle* 5, 327–335 (2006)
3. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. *ACM Journal of Computing Surveys* 38(4), 13 (2006)
4. Mukherjee, D., Ray, N., Acton, S.: Level set analysis for leukocyte detection and tracking. *IEEE Trans. on Image Processing* 13(4), 562–572 (2004)
5. Li, K., Miller, E., Weiss, L., Campbell, P., Kanade, T.: Online tracking of migrating and proliferating cells imaged with phase-contrast microscopy. In: *Proceedings of Intern. Works. on Comp. Vision and Pattern Recog.*, pp. 65–72 (June 2006)
6. Padfield, D., Rittscher, J., Thomas, N., Roysam, B.: Spatio-temporal cell cycle phase analysis using level sets and fast marching methods. In: *Proc. of first Workshop on Microscopic Image Analysis with Applications in Biology*, pp. 2–9 (2006)
7. Xie, J., Tsui, H.: Image segmentation based on maximum-likelihood estimation and optimum entropy-distribution (MLE-OED). *Pattern Recognition Letter* 25(10), 1133–1141 (2004)
8. Xu, C., Prince, J.L.: Snakes, shapes, and gradient vector flow. *IEEE Trans. on Image Processing* 7(3), 359–369 (1998)
9. Nocedal, J., Wright, S.J.: *Numerical Optimization*. Springer, Heidelberg (1999)
10. Dempster, A.: Maximum likelihood from incomplete data via the em algorithm. *J. Roy. Statist. Soc.* 29, 1–38 (1997)
11. Rosenfeld, A., Kak, A.C.: *Digital Picture Processing*. Academic Press, London (1982)
12. Comaniciu, D., Ramesh, V., Meer, P.: Real-time tracking of non-rigid objects using mean-shift. In: *Proc. of IEEE. Conf. Comp. Vision and Pattern Recog.*, vol. II, pp. 142–149 (June 2000)
13. Hopcroft, J., Karp, R.: An $n^{2.5}$ algorithm for maximum matchings in bipartite graphs. *SIAM Journal on Computing* 4(2), 225–230 (1979)
14. Garey, M.R., Johnson, D.S.: *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman, New York (January 1979)
15. Shafique, K., Shah, M.: A noniterative greedy algorithm for multiframe point correspondence. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 27, 51–65 (2005)
16. Veenman, C., Reinders, M., Backer, E.: Resolving motion correspondence for densely moving points. *IEEE Trans. on Pat. Anal. and Machine Intelligence* 23(1), 54–72 (2001)
17. Wright, S., Walia, B., Parkinson, J., Khan, S.: Differential activation of escherichia coli chemoreceptors by blue-light stimuli. *J. Bacteriol.* 188, 3962–3971 (2006)
18. Deb, S., Yeddanapudi, M., Pattipati, K., Bar-Shalom, Y.: A generalized s-d assignment algorithm for multisensor-multitarget state estimation. *IEEE Trans. Aerospace and Electronic Systems* 33(2), 523–538 (1997)
19. Poore, A.: Multidimensional assignments and multitarget tracking. In: *Proc. Partitioning Data Sets*, pp. 169–196 (1995)
20. Grundland, M., Dodgson, N.A.: Automatic contrast enhancement by histogram warping. In: *Proc. of Intern. Conf. Comp. Vision and Graphics*, pp. 22–24 (September 2004)
21. Collins, R.: Mean-shift blob tracking through scale space. In: *Proc. of IEEE. Conf. Comp. Vision and Pattern Recog.*, pp. 234–240 (June 2003)