A Matching Model Based on Earth Mover's Distance for Tracking *Myxococcus Xanthus*⋆

Jianxu Chen[1], Cameron W. Harvey[2], Mark S. Alber[2,1], and Danny Z. Chen[1]

[1] Department of Computer Science and Engineering, University of Notre Dame, USA
[2] Department of Appl. and Comput. Math and Stat., University of Notre Dame, USA

Abstract. Tracking the motion of *Myxococcus xanthus* is a crucial step for fundamental bacteria studies. Large number of bacterial cells involved, limited image resolution, and various cell behaviors (e.g., division) make tracking a highly challenging problem. A common strategy is to segment the cells first and associate detected cells into moving trajectories. However, known detection association algorithms that run in polynomial time are either ineffective to deal with particular cell behaviors or sensitive to segmentation errors. In this paper, we propose a polynomial time hierarchical approach for associating segmented cells, using a new Earth Mover's Distance (EMD) based matching model. Our method is able to track cell motion when cells may divide, leave/enter the image window, and the segmentation results may incur false alarm, detection lost, and falsely merged/split detections. We demonstrate it on tracking *M. xanthus*. Applied to error-prone segmented cells, our algorithm exhibits higher track purity and produces more complete trajectories, comparing to several state-of-the-art detection association algorithms.

1 Introduction

Myxococcus xanthus is a Gram-negative myxobacterium that has an elongated rod shape with rounded ends and moves by gliding on surfaces [4]. As an organism exhibiting sophisticated collective motions, *M. xanthus* is recognized as an important model for studying bacteria motility mechanism and bacteria swarming [11]. To gain insights into their motion patterns, quantitative study of moving trajectories of bacterial cells is needed. Due to the laboriousness of manually tracking large numbers of densely moving bacterial cells in long image sequences, reliable automated or semi-automated bacteria tracking algorithms become crucial for time-resolved quantitative analysis of bacteria swarming.

For motion tracking of *M. xanthus*, a robust scheme should be capable of detecting cell divisions and cells leaving/entering the image window. In practice, we face additional challenges. Halos in the images (e.g., Fig. 1) may lead to false alarms. Also, *M. xanthus* may appear in large-sized clusters. For clusters with cells touching tightly or temporarily out-of-focus, it is likely to result in track

⋆ This work was supported in part by NSF under Grant CCF-1217906 and by NIH under Grants 1R01-GM095959 and 1R01-GM100470.

P. Golland et al. (Eds.): MICCAI 2014, Part II, LNCS 8674, pp. 113–120, 2014.

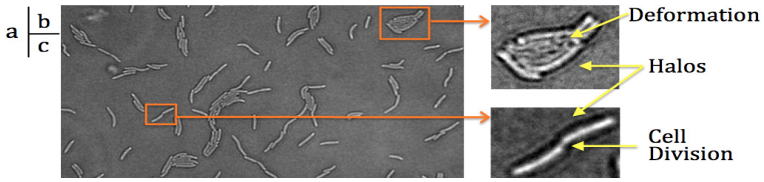

Fig. 1. (a) An image of *M. xanthus* swarms. (b) Halos may induce false alarms and shape deformation may result in track lost or falsely merged detections. (c) A cell is undergoing normal binary cell division. It will divide into two in the next few frames.

lost or falsely merged detections. Further, possible intensity degradation may cause false split, or even missing, detections of the tenuous rod shape cells.

Known cell tracking methods can be roughly classified into three categories [8]: probabilistic methods, contour evolution, and detection-based association (DBA). In general, probabilistic methods (e.g., JPDA [3]) can track fast moving cells, but cannot handle cell divisions well. Evolution based methods (e.g., level sets [6]) are effective in handling cell divisions, and even robust to local degradation of image quality. But, such methods are less effective when cells may tightly touch in clusters or when new cells can move into the image window. DBA methods (usually, cell segmentation followed by cell matching) are effective in handling the aforementioned cell behaviors [7,8,12]. But, known DBA methods suffer from their sensitivity to even small segmentation errors, which are hard to avoid (e.g., in cell clusters in Fig. 1), or high time-complexity [1,10].

Recently, a nesting Earth Mover's Distance (EMD) model for tracking evolving data clusters was proposed [5], which could be useful to cell tracking. EMD [9] is a function measuring the distance between distributions in a given feature space. But, in the nesting EMD model [5], one of the key steps was formulated as a binary integer problem, which is NP-complete.

In this paper, we extend the standard EMD metric to an EMD matching model, and develop a new DBA method based on it, i.e., a hierarchical scheme for associating detected cells. We show its application on tracking *M. xanthus*. Assume we have obtained error-prone segmentation results (say, by an automated segmentation algorithm). We take three main steps to extract cell trajectories: (1) Apply our EMD matching model to every two consecutive image frames to build a reliable frame-to-frame correspondence; (2) perform a global association on the unmatched cells (also formulated as and solved by the EMD matching model); (3) analyze the cell correspondence by track parsing to identify the actual trajectory of each cell. The entire process takes polynomial time in terms of the numbers of cells and frames. Evaluated by tracking *M. xanthus*, our method can identify more complete trajectories than several state-of-the-art association algorithms on segmentation results that contain various degrees of errors.

The essential difference between our method and other DBA methods is as follows. Most previous association algorithms are based on either some one-to-one matching, which is highly sensitive to segmentation errors, or formulated

as large-sized integer programming problems, which are in general NP-hard, to explicitly model cell behaviors and segmentation errors. In contrast, our EMD matching model, which is a transportation problem in essence, implicitly accommodates all cell behaviors and segmentation errors by utilizing multiple-to-one/one-to-multiple matchings in a natural way. Thus, our algorithm is capable of tackling all the aforementioned challenges and still runs in polynomial time.

2 Methodology

2.1 Earth Mover's Distance (EMD)

In computer vision, EMD is a metric for measuring the difference between two multidimensional distributions in the feature space [9]. Below, we sketch EMD.

The input of EMD is two *signatures*, $\mathcal{P} = \{(p_1, w_{p_1}), \ldots, (p_n, w_{p_n})\}$ and $\mathcal{Q} = \{(q_1, w_{q_1}), \ldots, (q_m, w_{q_m})\}$; p_i and q_j are application-dependent representatives (e.g., clusters of points) in the feature space; w_{p_i} and w_{q_j} are associated weights (e.g., the number of points in a cluster). A key component of EMD is the *ground distance* between p_i and q_j, denoted by D_{ij}. Intuitively, if viewing the weight as mass, then D_{ij} measures the cost for distributing one unit of mass from p_i to q_j, while EMD measures the smallest average cost for mass distribution from $\mathcal{P}$ to $\mathcal{Q}$. Let f_{ij} denote the mass flow from p_i to q_j. EMD is defined as

$$EMD(\mathcal{P}, \mathcal{Q}) = \frac{\min\limits_{f_{ij}} \sum\limits_{1 \le i \le n} \sum\limits_{1 \le j \le m} D_{ij} f_{ij}}{\sum\limits_{1 \le i \le n} \sum\limits_{1 \le j \le m} f_{ij}} \tag{1}$$

subject to

$$\sum_{1 \le j \le m} f_{ij} \le w_{p_i}, \forall 1 \le i \le n; \qquad \sum_{1 \le i \le n} f_{ij} \le w_{q_j}, \forall 1 \le j \le m \tag{2a}$$

$$\sum_{1 \le i \le n} \sum_{1 \le j \le m} f_{ij} = \min\Big(\sum_{1 \le i \le n} w_{p_i}, \sum_{1 \le j \le m} w_{q_j}\Big) \tag{2b}$$

$$f_{ij} \ge 0, \forall 1 \le i \le n, 1 \le j \le m \tag{2c}$$

The essence of EMD is a transportation problem, which can be solved in polynomial time [2]. In the context of transportation, p_i is a supplier and q_j is a consumer, and the weights are their capacities. Then, (2a) restricts each supplier (resp., consumer) not to send (resp., receive) more than its capacity; (2b) ensures that either all suppliers are exhausted or all consumers are satisfied; (2c) means flows can only move from suppliers to consumers.

2.2 EMD Based Matching Model and Hierarchical Association

In our EMD matching model, the goal is to find the optimal correspondence between two sets of cells, $\mathcal{P} = \{(p_1, w_{p_1}), \ldots, (p_n, w_{p_n})\}$ and $\mathcal{Q} = \{(q_1, w_{q_1}), \ldots, (q_m,$

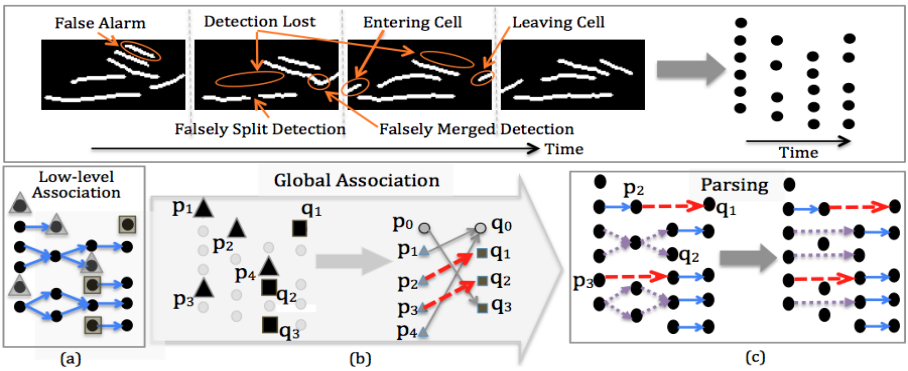

Fig. 2. Illustrating the hierarchical scheme. The top row gives detection results of four consecutive frames with errors (each detected cell is viewed as a vertex in a graph of four columns, one column per frame). The bottom row shows the effect of each step: (a) The correspondence of low-level association is shown by blue arrows; (b) using global association (p_0, q_0 are virtual cells) on the child-deficient cells ($\triangle$) and parent-deficient cells ($\square$), unmatched detections are linked optimally (red dashed arrows); (c) falsely merged/split trajectories are separated/combined (purple dot arrows) by track parsing.

$w_{q_m})\}$; p_i (resp., q_j) is a cell with w_{p_i} (resp., w_{q_j}) pixels in a 2D image. Two virtual cells p_0 and q_0 with very large weights, e.g., $w_{p_0} = w_{q_0} = \infty$, are also created. Defining the ground distance D_{ij} is crucial and depends on specific cell behaviors (see Section 2.4 for details). Intuitively, for $i, j \neq 0$, D_{ij} measures the matching cost of two detected cells. For $i \neq 0$ and $j = 0$ (resp., $i = 0$ and $j \neq 0$), D_{ij} is the cost for p_i leaving (resp., q_j entering) the image window, if p_i (resp., q_j) is within the vicinity of the image boundary; otherwise, D_{ij} is the cost for p_i (resp., q_j) matching to nothing, which means either p_i (resp., q_j) is a false positive detection or its corresponding object in $\mathcal{Q}$ (resp., $\mathcal{P}$) is missing. $D_{00} = \infty$, because flow between virtual cells is meaningless. By computing $EMD(\hat{\mathcal{P}}, \hat{\mathcal{Q}})$, with $\hat{\mathcal{P}} = (p_0, w_{p_0}) \cup \mathcal{P}$ and $\hat{\mathcal{Q}} = (q_0, w_{q_0}) \cup \mathcal{Q}$, we find the optimal cell correspondence between $\mathcal{P}$ and $\mathcal{Q}$, in the sense that the minimum EMD value means the smallest average matching cost and a large f_{ij} ($i \neq 0$, $j \neq 0$) implies a strong correspondence between p_i and q_j.

The low level of our hierarchical scheme (Fig. 2(a)) is to associate cells in every two consecutive frames of the input image sequence. Let $\mathcal{P}$ and $\mathcal{Q}$ be the two sets of cells in frames k and $k+1$, respectively. Then the above matching model can be applied naturally to associate cells in $\mathcal{P}$ and $\mathcal{Q}$. In the low-level association, we accept only reliable correspondence, i.e., $p_i \in \mathcal{P}$ is associated with $q_j \in \mathcal{Q}$ only if the f_{ij} value in the optimal solution is larger than 3/4 of the size of p_i or q_j and their matching cost, D_{ij}, is smaller than a threshold. For problems of large scales, we can divide each frame into overlapping subregions, and conduct the low level association in each subregion for computational efficiency.

The global association is applied to the low level association results (Fig. 2(b)). For a cell in frame k, its corresponding cells in frame $k+1$ (resp., $k-1$) are called

children (resp., *parents*). For each cell not in the last (resp., first) frame, if the total weight of its children (resp., parents), W_{sum}, and its own weight, W_{cell}, satisfy $|W_{sum} - W_{cell}| < 1/4 \cdot |W_{cell}|$, we call it *child-sufficient* (resp., *parent-sufficient*); otherwise, it is *child-deficient* (resp., *parent-deficient*). Let $\mathcal{P}_g$ (resp., $\mathcal{Q}_g$) be the set of all child-deficient (resp., parent-deficient) cells in the low level association results of the entire image sequence, and the weight of each such cell be $|W_{sum} - W_{cell}|$. Again, we apply our EMD matching model to find the optimal correspondence between $\mathcal{P}_g$ and $\mathcal{Q}_g$. In the optimal solution, the correspondence between a cell in $\mathcal{P}_g$ and a cell in $\mathcal{Q}_g$ actually links a child-deficient cell to a parent-deficient cell to reconnect a "broken" trajectory.

2.3 Track-Parsing

Our model can identify cell fusion or division. But, the merge/split events resulted from segmentation errors should be eliminated (Fig. 2(c)). We employ a track-parsing step to separate or combine falsely merged or split trajectories.

Suppose a trajectory splits into two at a cell p. If the split trajectories merge again within the next K frames, say at q, we treat it as a false splitt event (e.g., S1 in Fig. 3) and link p to q directly. If one branch of the split trajectories terminates within the next K frames, we treat this short branch trajectory as a false one caused by false positive detection (e.g., S2 in Fig. 3), and remove this branch. Otherwise, we consider it as a cell division (e.g., S3 in Fig. 3).

Suppose two detected cells p_1 and p_2 merge and form a single trajectory. If it splits into two, say at q_1 and q_2, within the next K frames, we treat it as a false merging event (e.g., M1 in Fig. 3) and apply the EMD matching model on $\{p_1, p_2\}$ and $\{q_1, q_2\}$ to connect them directly. If one of the merged trajectories is initiated just within the previous K frames, we consider it as a false trajectory (e.g., M2 in Fig. 3) and remove the short branch. Otherwise, we separate the merged trajectories by selecting the best correspondence, i.e. smallest matching cost, because cells cannot be fused in our problem (e.g., M3 in Fig. 3). For applications where objects can merge, M3 is treated as a potential fusion.

If a merge/split event involves more than two cells, we examine two of them each time. Because we only check K frames (we use $K = 5$, but it is not sensitive), forward of backward, at each merge/split event, the track-parsing step only takes linear time in terms of the number of merge and split events, which is small in practice and upper bounded by the number of cells times the number of frames.

2.4 Ground Distance

A key component of EMD is the ground distance D_{ij}. In our problem, we define D_{ij} specifically for *M. xanthus* motion. It hinges on the observation that *M. xanthus* can glide forward or backward, but rarely in directions normal to its body orientation. First, we represent each rod-shape cell by its centerline curve as a pixel sequence. (When tracking cells in other shapes, D_{ij} should be defined accordingly.) Let $A = (l_1, l_2, \ldots, l_a)$ and $B = (r_1, r_2, \ldots, r_b)$ be the centerline representations of two cells. Suppose $a \leq b$. A has four possible ways to move to

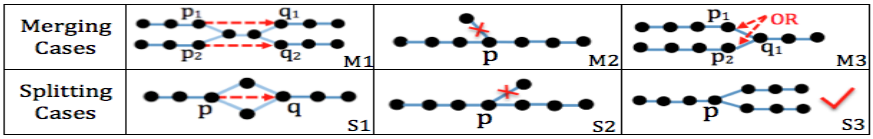

Fig. 3. Illustrating different possible cases of trajectory merging/splitting

the position of B (with two of their ends aligned together), which we represent by four permutations $\sigma_k \cdot (l_1, \ldots, l_a) = (r_{\sigma_k^{-1}(1)}, \ldots, r_{\sigma_k^{-1}(a)}), \forall 1 \leq k \leq 4$: $\sigma_1(i) = i$, $\sigma_2(i) = i + b - a$, $\sigma_3(i) = b - i + 1$, and $\sigma_4(i) = a - i + 1$, $\forall 1 \leq i \leq a$. Then, the ground distance between A and B, denoted by $D_{A,B}$, is defined as $D_{A,B} = \min_{1 \leq k \leq 4} D^{\sigma_k}_{A,B}$, where $D^{\sigma_k}_{A,B}$ is

$$D^{\sigma_k}_{A,B} = (\frac{1}{a}\sum_{i=1}^{a} \|\boldsymbol{l_i r_{g_k(i)}}\|) \cdot (1 - \cos^4(\frac{1}{2a}\sum_{i=1}^{a} \langle \boldsymbol{l_i r_{g_k(i)}}, \boldsymbol{l_i} \rangle + \langle \boldsymbol{l_i r_{g_k(i)}}, \boldsymbol{r_{g_k(i)}} \rangle))) \quad (3)$$

$\| \cdot \|$ is the vector norm. $g_k(i) = \sigma_k^{-1}(i)$ denotes the index of the corresponding pixel of l_i in B under permutation σ_k. $\boldsymbol{l_i r_{g_k(i)}}$ is the vector from l_i to $r_{g_k(i)}$, $\boldsymbol{l_i}$ is the tangent vector at l_i, $\langle \boldsymbol{l_i r_{g_k(i)}}, \boldsymbol{l_i} \rangle$ is the acute angle formed by the two vectors. The first term measures the average distance between l_i and $r_{g_k(i)}$; the second term reflects the average deviation from matched pixels to the body orientations of A and B. To reduce the computation, the ground distance is calculated only between feasible matching cell pairs (classified by a distance threshold).

Next, we define the ground distance between a cell and a virtual cell. $D_{i0} = C_{null}(p_i)$ if p_i is not in the vicinity of the image boundary; otherwise, $D_{i0} = C_e(p_i)$ (similarly for D_{0j}). Here, $C_{null}(\cdot)$ is set as the maximum ground distance between all feasible matches; $C_e(p_i) = \beta * (d_{min} + w_{p_i})$, where β is a parameter that acts like the second term in Equ. (3) (β is computed by assuming the cell orientation changes by $\pi/6$ on average; then $\beta = 1 - \cos^4(\pi/6) = 0.4$), and d_{min} is the minimum distance from any pixel on p_i to the image boundary. p_i is in the image boundary vicinity only if d_{min} is smaller than a distance threshold. For low frequency images as ours, this threshold should be relatively large (set to 10 pixels in our implementation). Note that $D_{00} = \infty$; otherwise, D_{ij} requires $\mathcal{O}(a)$ operations. In fact, $\mathcal{O}(a) \leq L * \mathcal{O}(1)$, where L is maximum cell length (about 40 pixels in our problem). Thus, D_{ij} can be computed in constant time.

3 Evaluation and Discussions

The data for evaluation are 16-bit grayscale image sequences obtained in experiments (e.g., Fig. 1). Seven image sequences were segmented by the method in [7] and used for tracking. Among them, the segmentation of the longest sequence SQ7 was manually examined and corrected by human experts. For the other six sequences, the segmentation errors, visually checked by comparing the segmented cells with the raw images, are summarized in Table 1. Regarding detection association, our algorithm (EMD) is compared with three state-of-the-art

Table 1. Evaluation results: Segmentation errors include false positive (FP), false negative (FN), false merge (FM), and false split (FS) detections. SQ7 is a manually corrected sequence. Column 2 (Cell No.) shows the number of trajectories that should be identified. Columns 8 to 11 show the numbers of completely detected trajectories by four methods, Local [7], Frame [8], Global [12], and our algorithm (EMD)

Sequence Info.			Segmentation Error (%)				Correct Trajectories			
No.	Cell No.	Frame No.	FP	FN	FM	FS	Local	Frame	Global	EMD
SQ1	58	75	0.57	15.5	3.0	5.62	23	14	18	**41**
SQ2	46	50	0.05	0.1	1.66	1.66	26	21	25	**35**
SQ3	51	85	0.03	1.56	1.14	1.48	7	2	17	**33**
SQ4	70	51	1.33	0.16	0.96	0.16	43	34	16	**57**
SQ5	50	40	0.43	1.19	1.11	1.02	30	22	29	**42**
SQ6	48	64	0.19	0.38	0.26	0.94	38	27	37	**47**
SQ7	142	100	0	0	0	0	121	108	96	**135**

methods that perform association in different levels: (i) A local method [7] (Local) performs neighborhood-based bipartite graph matching for each cell to find its corresponding cell within a certain range in the next frame; (ii) a frame level approach [8] (Frame) computes a minimum-cost flow to build correspondence between cells in every two consecutive frames; (iii) a global strategy [12] (Global) applies k-partite matching to associate cells in k consecutive frames.

Our evaluation was conducted in two parts. First, the manually corrected segmentation results (SQ7) were annotated by human experts and served as the ground truth of cell correspondence. Next, 1% to 10% of the cells in SQ7 are randomly removed to simulate false negative (FN) detections. In order to evaluate the robustness to false split (FS), 1% to 10% of the cells in SQ7 are randomly selected and cut into two at the middle of each cell. Track purity [1] is measured, i.e., the number of correct correspondence over the total number of correspondence built by a tracking algorithm. The results are in Fig. 4. Our algorithm shows the best overall performance in the sense that we have only less than 4% error in trajectories in the worst case (with FN = 10% and FS = 10%).

Second, we compared the tracking performance of the above four methods directly on error-prone segmentation results. A unique index number is assigned

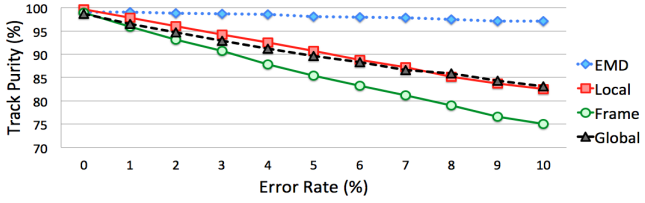

Fig. 4. Comparing the performance of different association methods in different degrees of false negative (FN) and false split (FS) segmentation errors. At an error rate k% on the x-axis, the y-axis value is the average of both the track purity of the tracking result with FN = k% and the track purity of the tracking result with FS = k%.

to each identified trajectory and labeled in the middle of the corresponding cells in the raw images. Then, the tracking results are visually examined. Any incomplete trajectory or falsely linked trajectory is treated as an error. The results are summarized in Table 1. In practice, even if only one detected cell is an error in each frame, errors can impact different trajectories in different frames. For association methods sensitive to segmentation errors, tiny error in cell segmentation may lead to big errors in computed trajectories (falsely linked or broken). Comparing to the other three methods, our algorithm can track more cells correctly in the presence of different degrees of segmentation errors.

4 Conclusions

We presented a hierarchical approach based on a new EMD matching model for associating segmented cells, which is a key process for detection-based association methods for cell tracking. Our algorithm takes polynomial time in terms of the numbers of cells and frames. Evaluated on motion tracking of *M. xanthus*, we showed that our method exhibits higher robustness to segmentation errors than several known association methods. For error-prone segmentation results, our method can identify more complete trajectories than these methods.

References

1. Bise, R., Yin, Z., Kanade, T.: Reliable cell tracking by global data association. In: ISBI, pp. 1004–1010 (2011)
2. Hillier, F., Lieberman, G.: Introduction to Operations Research, 8th edn. McGraw-Hill (2010)
3. Kachouie, N., Fieguth, P.: Extended-Hungarian-JPDA: Exact single-frame stem cell tracking. IEEE Trans. on Biomedical Eng. 54(11), 2011–2019 (2007)
4. Kaiser, D.: Social gliding is correlated with the presence of pili in *Myxococcus xanthus*. Proc. of the Nat. Acad. of Sci. 76(11), 5952–5956 (1979)
5. Kremer, H., Gunnemann, S., Wollwage, S., Seidl, T.: Nesting the earth mover's distance for effective cluster tracing. In: ICSSDM (2013)
6. Li, K., Chen, M., Kanade, T.: Cell population tracking and lineage construction with spatiotemporal context. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 295–302. Springer, Heidelberg (2007)
7. Liu, X., Harvey, C.W., Wang, H., Alber, M.S., Chen, D.Z.: Detecting and tracking motion of *Myxococcus xanthus* bacteria in swarms. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 373–380. Springer, Heidelberg (2012)
8. Padfield, D., Rittscher, J., Roysam, B.: Coupled minimum-cost flow cell tracking for high-throughput quantitative analysis. Med. Image Anal. 15(4), 650–668 (2011)
9. Rubner, Y., Tomasi, C., Guibas, L.: A metric for distributions with applications to image databases. In: ICCV, pp. 59–66 (1998)
10. Schiegg, M., Hanslovsky, P., Kausler, B.X., Hufnagel, L., Hamprecht, F.: Conservation tracking. In: ICCV, pp. 2928–2935 (2013)
11. Wu, Y., Kaiser, D., Jiang, Y., Alber, M.: Periodic reversal of direction allows myxobacteria to swarm. Proc. of the Nat. Acad. of Sci. 106(4), 1222–1227 (2009)
12. Xie, J., Khan, S., Shah, M.: Automatic tracking of *Escherichia coli* bacteria. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 824–832. Springer, Heidelberg (2008)