A Comparative Method for Analysing Toponome Image Stacks

Andrei Barysenka[1,*], Andreas W. M. Dress[1] and Walter Schubert[2]

[1] *Department of Combinatorics and Geometry, CAS-MPG Partner Institute and Key Lab for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China and Max Planck Institute for Mathematics in the Sciences, D-04103 Leipzig, Germany.*

[2] *Molecular Pattern Recognition Research Group Medical Faculty, Otto-von-Guericke University, Leipziger Str. 44 Magdeburg, 39120 Germany.*

Abstract. We present a technique to find threshold values that allows the user to separate signal from noise in fluorescence grey-level images. It can be classified as a purely *comparative* method based upon the amount of "Mutual Information" between two or more florescence images, and we apply it to stacks of such images produced using the newly-developed MELK technology. Our results are compared to results obtained by another research group using a quite different (completely independent and more technology-based) approach; and also to results obtained using *Otsu's Thresholding Method*, yet another completely distinct approach invented to separate foreground and background in a grey-level image, based on minimising "intra-class variance" [9, 10]. The remarkably good agreement found suggests that our proposed comparative information based method not only accounts for the biological mechanisms governing cellular protein networks very well, but also (and probably much more importantly) shows that cells actually organise the spatial structure of their protein networks in a highly non-random fashion as might be expected – and thereby try to optimise their "mutual information content", and thus most probably their efficiency.

Key words: MELK technology, Kullback-Leibler distance, threshold values, numerical optimization.

1. Introduction

Given a finite collection $\mathscr{C}$ of observations, each observation $c \in \mathscr{C}$ being represented by a vector $\varphi(c) = (c_1, \cdots, c_n) \in \mathbb{R}^n$, there are many areas – from sensing, security, and data mining to biology and medicine – where a pre-processing step assigning a $0, 1$-vector

---

*Corresponding author. *Email addresses:* `aboris@picb.ac.cn` (A. Barysenka), `andreas@picb.ac.cn` (A. W. M. Dress), `walter.schubert@med.ovgu.de` (W. Schubert)

$\chi(c) = \bar{c} = (\bar{c}_1, \cdots, \bar{c}_n)$ to each observation $c \in \mathscr{C}$ can be a great advantage. Clearly, it is appropriate that this assignment be *monotone* (i.e. $c, d \in \mathscr{C}$, $i \in \{1, \cdots, n\}$), that $c_i \leq d_i$ should always imply $\bar{c}_i \leq \bar{d}_i$, and that it should maximise the "surprise" (or the "mutual information content") that can be found in the data, following the "surprisology" paradigm proposed in [14].

The monotonicity requirement is easily met, by choosing a threshold or 'cut-off' vector $\mathbf{t} = (t_1, \cdots, t_n) \in \mathbb{R}^n$ that allows us to define a monotone map

$$\psi_{\mathbf{t}} : \mathbb{R}^n \to \{0, 1\}^n : (x_1, \cdots, x_n) \mapsto (\bar{x}_1, \cdots, \bar{x}_n)$$

by putting

$$\bar{x}_i := \begin{cases} 1 & \text{if } x_i \geq t_i, \\ 0 & \text{otherwise} \end{cases}$$

for all $i = 1, \cdots, n$. When composed with φ, this yields the required $\{0, 1\}$ vector $\bar{\mathbf{c}} = \psi_{\mathbf{t}}(c_1, \cdots, c_n) = \psi_{\mathbf{t}}(\varphi(c))$ for each observation $c \in \mathscr{C}$.

We follow the basic exposition already outlined in [2], by assuming that the mutual independence of the observed values $c_i, i \in \{1, \cdots, n\}$ of any observation $c \in \mathscr{C}$ (i.e. over all observations in $\mathscr{C}$) would *not* yield a surprise. Then one can measure the 'amount of surprise' encountered when associating a $\{0, 1\}$ vector $\boldsymbol{\chi}(c) = (\chi_1(c), \cdots, \chi_n(c))$ to each $c \in \mathscr{C}$, by comparing the *observed* probability distribution p_χ defined for each $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n) \in \{0, 1\}^n$ by

$$p_\chi(\boldsymbol{\varepsilon}) := \frac{\#\{c \in \mathscr{C} | \boldsymbol{\chi}(c) = \boldsymbol{\varepsilon}\}}{\#\mathscr{C}} \tag{1.1}$$

with the *expected* probability distribution q_χ defined by

$$q_\chi(\boldsymbol{\varepsilon}) := \prod_{i=1}^{n} \frac{\#\{c \in \mathscr{C} | \chi_i(c) = \varepsilon_i\}}{\#\mathscr{C}}. \tag{1.2}$$

Given the independence of the coordinates, note that q_χ would essentially coincide with p_χ. Consequently, noting that $\boldsymbol{\varepsilon} \in \{0, 1\}^n$ and $q_\chi(\boldsymbol{\varepsilon}) = 0$ always implies $p_\chi(\boldsymbol{\varepsilon}) = 0$, the well-known *Kullback-Leibler Divergence* or *Mutual-Information Function*

$$MI(q_\chi \to p_\chi) := \sum_{\boldsymbol{\varepsilon} \in \{0,1\}^n} p_\chi(\boldsymbol{\varepsilon}) \log \left(\frac{p_\chi(\boldsymbol{\varepsilon})}{q_\chi(\boldsymbol{\varepsilon})} \right) \tag{1.3}$$

(with $p_\chi(\boldsymbol{\varepsilon}) \ln(p_\chi(\boldsymbol{\varepsilon})/q_\chi(\boldsymbol{\varepsilon})) := 0$ for $p_\chi(\boldsymbol{\varepsilon}) = 0$) appears to be a good measure of that 'amount of surprise' encountered by the map $\boldsymbol{\chi} : \mathscr{C} \to \{0, 1\}^n$ (see e.g., [11], for its various virtues). Thus to each cut-off vector $\mathbf{t} \in \mathbb{R}^n$ one can associate its surprise value $surp(\mathbf{t}) = surp_{(\mathscr{C}, \varphi)}(\mathbf{t})$ defined by

$$\text{surp}(\mathbf{t}) := MI(q_{\psi_{\mathbf{t}} \circ \varphi} \to p_{\psi_{\mathbf{t}} \circ \varphi}), \tag{1.4}$$

and declare a cut-off vector $\mathbf{t}$ to be $(\mathscr{C}, \varphi)$-optimal if

$$\text{surp}(\mathbf{t}) \geq \text{surp}(\mathbf{t}')$$

holds for all $\mathbf{t}' \in \mathbb{R}^n$, and to be locally $(\mathscr{C}, \boldsymbol{\varphi})$-optimal if the above inequality holds for all $\mathbf{t}' \in \mathbb{R}^n$ in some sufficiently large neighbourhood $N(\mathbf{t})$ of $\mathbf{t}$.

We have designed and implemented a simple "greedy" algorithm that can quickly find locally $(\mathscr{C}, \boldsymbol{\varphi})$-optimal cut-off vectors $\mathbf{t}$ for n up to 10 or 12 and up to a million observations (the cardinality of $\mathscr{C}$). Many applications can be envisaged in other areas such as sensing and security, but here we apply our software to analyse multi-image data previously obtained using the Multi-Epitope Fluorescence Microscopy technology MELK discussed in [14] – cf. also [5, 12, 13]. Some details of this technology and tasks associated with the data so obtained are discussed in 2.1 below – cf. also [15, 16].

This particular application puts our procedure to a rather stringent test. Using their specific technology-based knowledge, scientists working with MELK have designed a completely different technology-specific approach to define a cut-off vector $\mathbf{t}$ for their multi-image data – and we find that both approaches yield essentially the same cut-off parameters, as discussed in detail in 2.3. Moreover, this is further evidence that the spatial distribution of protein co-localization patterns observed via MELK technology is highly non-random, which (whether explicitly already identified or not) must provide some quite significant information.

In the next section, we present the principal example for which our method was originally developed and the results that were obtained in this way, and then discuss some more general relevant aspects in the concluding section that follows.

2. Medical Image Segmentation

We first show that our approach can be used successfully to determine *threshold values* for separating the presence or absence of proteins in a stack of fluorescence images describing a spatial distribution of proteins across a biological object – e.g. a slice of nervous tissue or a sample of blood cells. More specifically, we apply our method to stacks of fluorescence images, and find threshold values that are almost identical to those found using completely independent methods based on technological and biological aspects of those images.

2.1. The basis of MELK technology

One of the most important aspects of cellular protein networks is the spatial distribution of the proteins across cell compartments (membranes, nuclei, mitochondria, organelles, etc.). Consequently, for some cellular function (e.g cell migration) to occur, a cell not only has to synthesise the necessary amounts of specific proteins but also distribute them internally and in its environment in a specific way required for that cell function. Conventional proteomics-profiling tools based on *homogenising* cell samples do not (almost by definition) provide any information on this important aspect of the molecular processes taking place in a cell — an aspect that is now well understood to be closely related to both the normal and the abnormal functioning of the cell. To provide such information, the new multi-
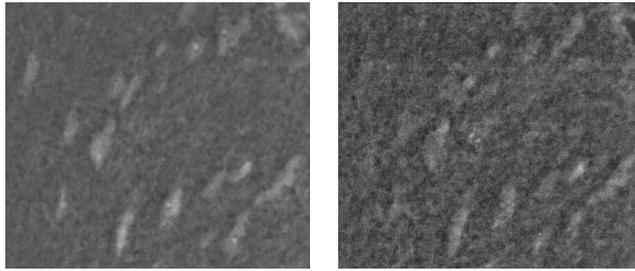
Figure 1: Two fluorescence images showing two views of the same specimen (a slice of nervous tissue), where each image displays the spatial distribution of one of two different proteins.

parameter fluorescence-microscopy technique called MELK[†] was developed by one of us, with coworkers at Magdeburg University [7, 12–14, 20]. This technique produces a stack of intensity images of a particular biological object such as a slice of nervous tissue, where each image in the stack corresponds to a particular extracellular or intracellular protein (or some other biologically relevant molecule of interest). Two typical grey-level images are shown in Fig. 1, and further biological applications of this technique are discussed elsewhere [3, 17–19, 21].

One of the most basic questions arising in this context is how the various observed fluorescence signals relate to the presence or absence of the corresponding proteins. It is clear that *high* intensity values of a fluorescence signal at a particular spot suggest a high protein concentration there (the corresponding protein is clearly present), whereas *low* intensity values indicate low or no such protein concentration — since due to 'noise', some non-zero intensity will be measured even where that concentration is zero (so some low intensity values should be interpreted as that protein being absent). However, to decide exactly what 'high' and 'low' should mean in this context, one needs to invoke a threshold that separates 'high' intensities from 'low' values, individually for each image in the given stack. In the next subsection, we describe a method for determining suitably optimal threshold values.

2.2. The multi-information function

Let us assume we have a stack of n fluorescence images of size N, where the term *size* refers to the number of pixels in the images under consideration. (In a typical image, we have $N = 512 \times 512$ pixels.) Each image shows the spatial distribution of one of n distinct fluorescent markers on the given tissue sample (see Fig. 1). These data can be regarded as a stack of *arrays* of integers (intensity values), lying in the interval $\{0, \cdots, 255\}$.

For $i = 1, \cdots, N$, let $I_j(i)$ denote the intensity value at the pixel i in the j-th image, and $\mathbf{I}(i) := \left(I_j(i)\right)_{j=1,\cdots,n}$ the corresponding n-dimensional integer vector of intensities; and for each $t = 0, \cdots, 255$, let $N_j(t)$ denote the number of pixels within the j-th image having
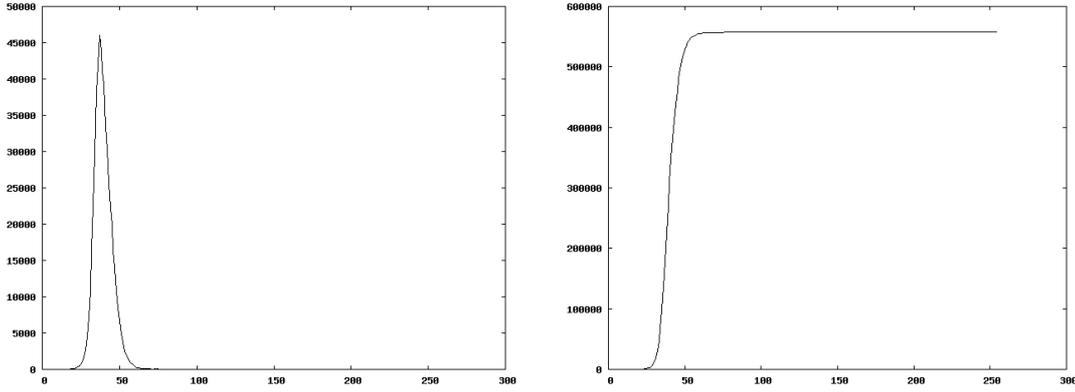

[†]*Multi-Epitop Liganden Kartographie*

Figure 2: Histogram of a fluorescence image (left) and the corresponding rank function $\mathrm{rank}(I)$ (right).

the intensity value t. Further, let

$$\mathrm{rank}_j(t) := \#\{i \,|\, I_j(i) < t\} = N_j(0) + \cdots + N_j(t-1) \qquad (2.1)$$

denote the number of pixels in the j-th image with intensities less than t (see Fig. 2).

If the fluorescence intensities of different fluorescent markers were independent of each other, for any given family

$$\mathbf{t} := (t_1, \cdots, t_n)$$

of numbers in $\{0, \cdots, 255\}$ one may expect the proportion of pixels with intensity values below t_j for *all* $j = 1, \cdots, n$ should coincide with the product $\prod_{j=1,\cdots,n} \mathrm{rank}_j(t_j)/N$ over all j of the corresponding proportions $\mathrm{rank}_j(t_j)/N$ of pixels with intensity values below t_j (for each individual image j). More generally, given any *sign vector* $\boldsymbol{\varepsilon} = (\varepsilon_j)_{j=1,\cdots,n} \in \{\pm 1\}^n$, total independence would imply that the proportion $p(\mathbf{t}, \boldsymbol{\varepsilon})$ of pixels with intensities

- below t_j for all $j = 1, \cdots, n$ with $\varepsilon_j = -1$ and

- equal to or above t_j for all other j

should coincide with the product

$$q(\mathbf{t}, \boldsymbol{\varepsilon}) := \prod_{\varepsilon_j = -1} \frac{\mathrm{rank}_j(t_j)}{N} \prod_{\varepsilon_j = 1} \frac{N - \mathrm{rank}_j(t_j)}{N} = \prod_{j=1,\cdots,n} \frac{\frac{\varepsilon_j + 1}{2} N - \varepsilon_j \mathrm{rank}_j(t_j)}{N} \,,$$

over all j, where the corresponding proportions of pixels of $\mathrm{rank}_j(t_j)/N$ or $(N - \mathrm{rank}_j(t_j))/N$ are respectively with intensity values below t_j for each j with $\varepsilon_j = -1$ and equal to or above t_j for each j with $\varepsilon_j = 1$. Thus each $\mathbf{t}$ defines two discrete probability distributions

$$p = p_{\mathbf{t}} := (p(\mathbf{t}, \boldsymbol{\varepsilon}) : \boldsymbol{\varepsilon} \in \{\pm 1\}^n)$$

and

$$q = q_{\mathbf{t}} := (q(\mathbf{t}, \boldsymbol{\varepsilon}) : \boldsymbol{\varepsilon} \in \{\pm 1\}^n)$$

on the set $\{\pm 1\}^n$ of sign vectors. The distribution q corresponds to total independence of the fluorescence signals in the different images – whereas p describes the actual data-point distribution among the various 'cells' defined by $\mathbf{t}$ and parametrised by the sign vectors $\boldsymbol{\varepsilon}$ in $\{\pm 1\}^n$, in the space $\{0, \cdots, 255\}^n \subseteq \mathbb{R}^n$ of all potential intensity vectors. If p and q were to coincide for all $\mathbf{t}$, studying the n signals simultaneously would not be of much interest. Consequently, by using the vector $\mathbf{t}$ as a vector of potential threshold values to decide on the presence or absence of proteins in the n images, one would maximise the information provided by the stack of images if $\mathbf{t}$ is chosen to maximise the distance between $p_\mathbf{t}$ and $q_\mathbf{t}$ measured, following the standard traditional *Kullback-Leibler Distance* $[4, 6]$

$$MI(\mathbf{t}) := MI(p(\mathbf{t}), q(\mathbf{t})) := \sum_{\boldsymbol{\varepsilon}} p(\mathbf{t}, \boldsymbol{\varepsilon}) \ln \frac{p(\mathbf{t}, \boldsymbol{\varepsilon})}{q(\mathbf{t}, \boldsymbol{\varepsilon})} \tag{2.2}$$

between $p(\mathbf{t})$ and $q(\mathbf{t})$, also called the Multi-Information Function $[22, 23]$. Maximising $MI(\mathbf{t})$ with respect to $\mathbf{t}$ provides our thresholds for deciding on the presence or absence of proteins in the n images - i.e. by maximising the 'mutual information content' that can be extracted from these images, using a threshold-based *presence/no-presence* decision process.

The numerical maximisation of $MI(\mathbf{t})$ was performed using a "greedy" approach. We first choose an initial threshold vector $\mathbf{t}_{start}$ randomly and then, starting with the axis $j = 1$, try to 'move' t_j until a local maximum of the function $MI(\mathbf{t})$ is reached along this particular axis. We then take the next axis $j = 2$ and perform the same search along it. After doing so for each of n axes, we start again with $j = 1$ and repeat these steps, until we find a vector $\mathbf{t}_{end}$ such that there is no remaining index $j \in \{1, \cdots, n\}$ where a variation could lead to further improvement, given the other coordinate values. On completing the above procedure, we reach a relative local maximum of the cost function $MI(\mathbf{t})$. Obviously, this sub-optimal threshold vector $\mathbf{t}_{end}$ may not provide the global maximum of the function $MI(\mathbf{t})$. We therefore repeated the algorithm several times with different starting vectors $\mathbf{t}_{start}$ (typically from 20 to 50), before choosing the $\mathbf{t}_{end}$ vector that gave the highest value of $MI(\mathbf{t})$.

To examine the presence of local maxima more carefully, we ran our algorithm 100 times – each time starting from a randomly chosen starting vector $\mathbf{t}_{start}$ – and counted the number of different local maxima and the corresponding "hits" (i.e. the number of runs our algorithm ended at a particular relative local maximum). The result of our simulation for the stack sb is shown in Fig. 3. The values of the cost function $MI(\mathbf{t})$ at the various relative local maxima are shown on the horizontal axis and the "hit" numbers for each on the vertical axis, and it appears that the largest of those relative local maxima has the largest "hit" number – which is in fact most likely the global maximum. (There also appear to be a number of other local maxima that are much less likely to be found.) We believe that the chance of finding the global optimum in this iterative way is high, given the almost convex shape of the scoring function implied by Fig. 5, which illustrates the $MI(\mathbf{t})$ for real fluorescence data corresponding to just two fluorescence images. However, a suitable theoretical analysis of the procedure and its expected success rate has not yet been carried out.
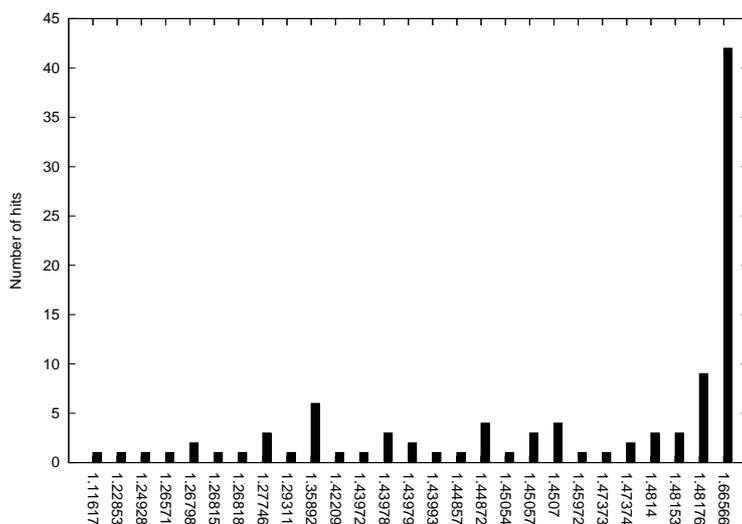
Figure 3: Numbers of "hits" for different local maxima for stack sb.
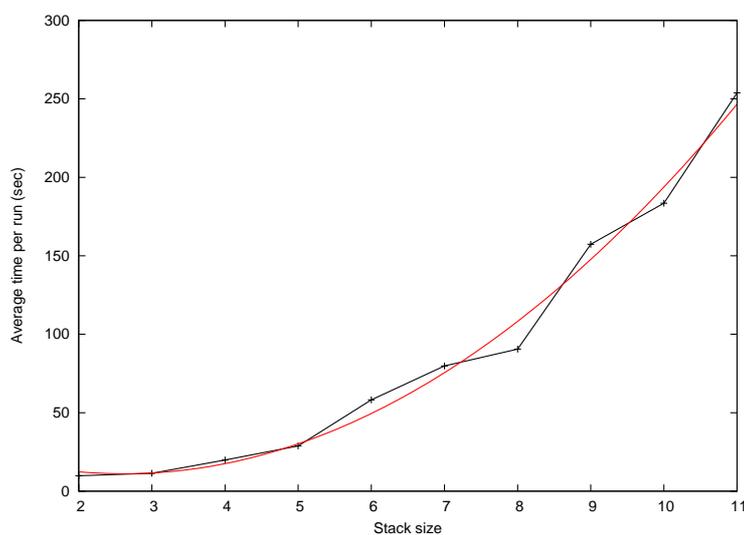
Figure 4: Average running time per run for substacks of size 2...11 from the stack sc (black) and the fitting quadratic curve (red).

Another rather important practical issue is the running time of our algorithm. To estimate this, we have run the algorithm on various substack sizes from the stack sc, with results as shown in Fig. 4. The running time seems to depend quadratically upon the substack size – cf. the fitted red curve. As might be expected, more time is necessary to complete the search for larger stacks (where the search space is so much larger).

In order to obtain the optimal threshold value for a given fixed protein i, hitherto we have proceeded to analyse whole image stacks. However, we could analyse some (or

Table 1: Threshold values found by analysing different pairs of images from the stack sa. The "all-in-one-go" thresholds are shown in bold.

(j,i)	1	2	3	4	5	6	7	8	9	10	11
1	**87**	87	87	87	85	88	86	90	93	90	81
2	67	**67**	66	69	68	67	68	66	80	79	86
3	101	100	**100**	99	98	101	100	101	109	107	92
4	96	98	96	**97**	97	97	97	95	107	107	116
5	55	60	56	60	**59**	58	56	59	79	73	92
6	92	91	91	92	91	**91**	91	96	106	99	76
7	59	62	60	61	59	61	**60**	61	71	64	79
8	63	62	62	62	62	63	62	**62**	63	63	59
9	11	12	11	12	12	11	11	11	**12**	12	62
10	61	66	63	66	64	62	60	60	67	**60**	72
11	78	56	77	54	57	85	54	61	56	54	**59**

even all) substacks $S_j := \{i_1^j, i_2^j, \cdots, j, \cdots i_k^j\}$ containing the $j-$th image to find the optimal threshold values $T_{S_j}(j)$ corresponding to those substacks and image j, but obviously these values do not have to coincide. It therefore seemed of interest to investigate how the choice of a particular substack S_j affects the optimal threshold for the $j-$th image, but we found the variability of the resulting $T_{S_j}(j)$ to be rather small – cf. Table 1, showing the optimal thresholds for each of the 11 images and each substack of size 2 for the stack sa. The nondiagonal $(j,i)-$th element in Table 1 gives the optimal threshold for the protein j and substack $\{i,j\}$. Except for the last three columns that tend be persistently higher than average, and the last row pertaining to Protein 11 that now seems to play a slightly erratic role (still to be investigated more closely from a biological viewpoint), the values in each row are quite close to each other and also to the value obtained from the whole stack – cf. the diagonal elements of Table 1, and also Table 2.

Thus one may speed up the algorithm considerably by choosing one reference image in a given stack and then determining the optimal threshold values for every other image in the stack relative to that reference image – and if the threshold values obtained for the reference image roughly coincide, one can safely take the obtained reference-image related threshold values as a good approximation of the de facto optima for further steps in image analysis. On the other hand, a clear disagreement could be a useful hint that there may be some interesting relationships between the proteins or other bio-molecules involved.

From the identity

$$\text{rank}_j(t) = \#\{i|I_j(i) < t\} = \#\{i|\text{rank}_j(I_j(i)) < \text{rank}_j(t)\}$$

it is notable that the value of $MI(\mathbf{t})$ depends only on the rank vector

$$\text{rank}(\mathbf{t}) := \left(\text{rank}_j(t_j)\right)_{j=1,\cdots,n}$$

associated with $\mathbf{t}$. This suggests *normalising* the map MI, say by introducing the associated real-valued map from the unit hypercube $[0,1]^n$ into $\mathbb{R}$ defined by mapping each vector
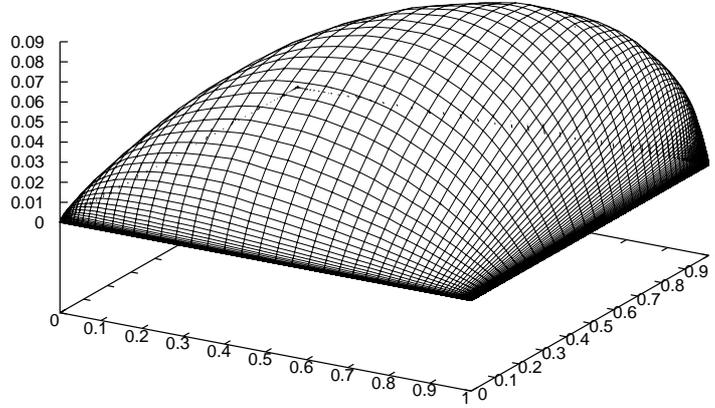
Figure 5: The function $MI(\mathbf{t})$ for a case with $n = 2$.

$\boldsymbol{\rho} = (\rho_1, \cdots, \rho_n) \in [0,1]^n$ onto the value that the map MI attains at the intensity vector $\mathbf{t} = (t_1, \cdots, t_n) = \mathbf{t}(\vec{\rho})$ for which

$$t_j := \min(t \in \mathbb{Z} : \text{rank}_j(t) \geq N\rho_j)$$

holds.[‡] At risk of abuse of notation, one may denote the resulting real-valued map defined on $[0,1]^n$ also by MI, which are the functions (smoothed at their discontinuities) to which the Figs. 5 and 7 below refer.

2.3. Results

Our method was applied to a data set consisting of 7 image stacks dubbed **sa** to **sg**, each containing 11 fluorescence images (Fig. 1 shows two typical images from this data set). As explained in 2.2, we determined optimal threshold values for each of the 7 stacks. Simultaneously, a group of scientists at Magdeburg University determined an alternative set of threshold values using methods for estimating plausible thresholds based on technological and biological considerations (cf. Tables 2 and 3).

It is notable that the sets of threshold values are quite similar, except for the 9th molecular species where the threshold values differ substantially in all seven stacks, which can be explained from the actual fluorescence images. In each data stack under investigation, the image corresponding to the 9th molecular species differs significantly from most of the images in the particular stack (cf. Fig. 6), being a rather 'black-and-white' than a grey-value

---

[‡]Clearly, this procedure is closely related to the *copula construction* discussed in [8] – cf. also [1] for further applications of this concept in image analysis.

Table 2: Values of thresholds for stacks **sa** − **sd** found by our *MI*-based method, followed by those determined by technology-based methods and by Otsu's method (successive data in parentheses).

Molecular	Thresholds			
species #	sa	sb	sc	sd
1	87 (88/86)	55 (59/55)	114 (112/124)	42 (44/69)
2	67 (64/68)	97 (99/96)	53 (58/53)	128 (141/130)
3	100 (98/98)	152 (157/151)	64 (75/64)	60 (61/61)
4	97 (96/102)	90 (93/92)	22 (25/22)	94 (94/94)
5	59 (55/68)	109 (114/111)	63 (68/64)	46 (48/48)
6	91 (83/87)	75 (78/76)	37 (37/38)	61 (63/63)
7	60 (62/59)	82 (86/83)	79 (86/79)	103 (111/104)
8	62 (66/61)	72 (76/72)	47 (50/47)	77 (72/77)
9	12 (50/54)	7 (57/51)	38 (37/62)	14 (55/81)
10	60 (61/57)	107 (107/110)	69 (77/69)	46 (51/45)
11	59 (66/79)	73 (73/78)	56 (66/75)	56 (54/67)

Table 3: Values of thresholds for stacks **se** − **sg**, found by our *MI*-based method, together with those determined by earlier technology-based methods and by Otsu's method (successive data in parentheses).

Molecular	Thresholds		
species #	se	sf	sg
1	110 (105/114)	48 (60/69)	108 (109/107)
2	71 (70/70)	62 (65/62)	59 (57/59)
3	81 (77/82)	41 (46/42)	105 (106/105)
4	61 (58/65)	89 (94/89)	140 (142/144)
5	122 (122/124)	67 (73/70)	117 (115/118)
6	51 (52/58)	43 (47/43)	109 (105/112)
7	75 (74/78)	61 (63/62)	158 (165/158)
8	37 (36/38)	74 (80/75)	85 (83/85)
9	39 (47/47)	15 (59/63)	33 (53/68)
10	129 (131/131)	37 (41/36)	136 (144/139)
11	83 (77/105)	41 (49/59)	119 (116/120)

image. As the optimal thresholds are obtained by maximising $MI(\mathbf{t})$, it is helpful to look at the function $MI(\mathbf{t})$ (see Fig. 5), where for each $i = 1, \cdots, 11$ we consider the *i-section* of $MI(\mathbf{t})$ through the maximum $T^* = (t_1^*, \cdots, t_{11}^*)$ of MI obtained using our algorithm - i.e. the graph of the function $f_i : t \mapsto MI(t_1^*, \cdots, t_{i-1}^*, t, t_{i+1}^*, \cdots, t_{11}^*)$. Three such sections for the **sb** stack are shown in Fig. 7. It can be seen immediately that the behaviour of $MI(\mathbf{t})$ along the 9-th axis is rather different from that along the other axes. Compared with other sections, the 9-section is rather flat over a long interval including t_9^*, corresponding to the intensities of the 9th molecular species being either rather high or rather low. Thus many threshold values lead to more or less the same decision regarding the presence or absence
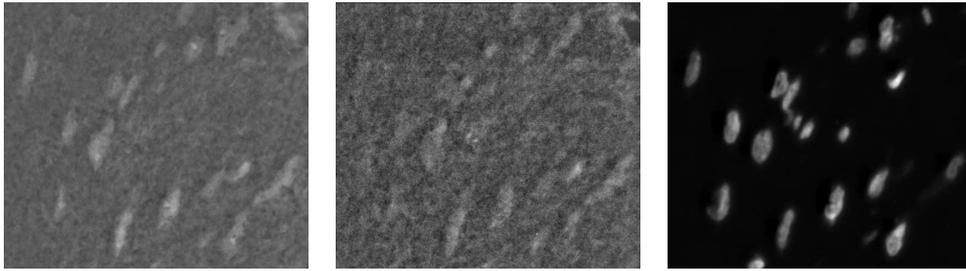
Figure 6: The fluorescence images corresponding to the molecular species 4 (left), 6, (center) and 9 (right) from the stack **sa**, where each image shows the same piece of nervous tissue.
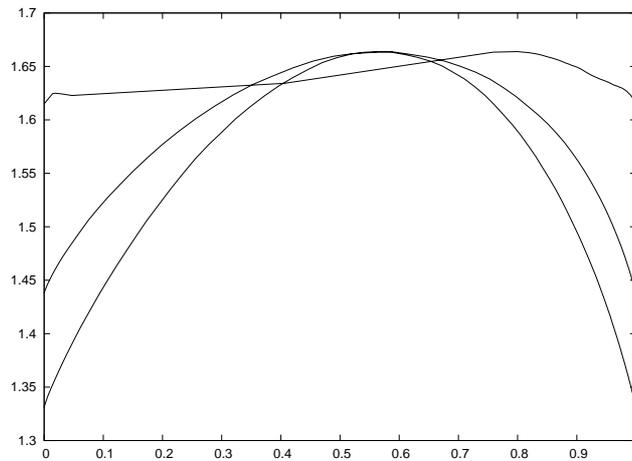

Figure 7: The maps f_1, f_4, and f_9 (relative to the normalised coordinates introduced above). The flat curve corresponds to the 9-section f_9, whereas the 1-section and the 4-section are each given by rather bell-shaped curves.

of that species, so it makes little difference whether the optimal threshold value $t_9 = 7$ or $t_9 = 57$ (cf. Table 2) is chosen, whether for the 'black-and-white' image resulting from a chosen threshold value (cf. Fig. 8) or for the map MI that is to be optimised. Actually, the 9th molecular species is not even a protein, but DNA - i.e. the 9th images indicate the presence or absence of mostly nuclear chromosomes (not a protein), so it is not surprising that these images have a rather different character than the others. Moreover, it is somewhat comforting to realise (and rather a corroboration than a refutation of our method) that the distinction is detected by our MI-based method, resulting in a flat rather than a smoothly curved shape.

For every image in the seven image stacks, our results are also compared in Tables 2 and 3 with those obtained by applying Otsu's well known thresholding method [9, 10], another completely distinct approach that simply tries to separate "foreground" and "background" in a gray-level image by minimizing "intra-class variance". With the exception of the special 9-th species case, once again the results agree closely. Thus for the given data, a simple

Table 4: Time requirements for Otsu's algorithm, for different stacks.

Stack	sa	sb	sc	sd	se	sf	sg
Time (sec)	0.60	0.84	0.72	0.64	0.62	0.76	0.64

Figure 8: Binary images corresponding to the 9th molecular species from the stack **sb**, with threshold value $t_9 = 7$ (left) and $t_9 = 57$ (right).

"background" and "foreground" separation apparently also identifies the fluorescent signal quite well, producing image stacks with high mutual information content.

The running times for Otsu's algorithm presented in Table 4 are much less than for our approach, shown in Fig. 4. This is not surprising, since Otsu's algorithm works with only one image at a time whereas ours is designed to process whole stacks of images. (Even for just 2 images, our search space is much larger than in Otsu's case.) However, it seems to be a biologically intriguing hint that both algorithms lead to roughly the same threshold values in most of the cases considered by us so far - i.e. the biological mechanisms controlling the spatial distribution of proteins in living cells are such that cells may actually try to increase the efficiency of their protein networks by optimising the "mutual information content" of that distribution.

3. Conclusion

Our method of simultaneously choosing individual threshold values for all images in a given stack of grey-value images is based on maximising the "amount" of mutual information between different random variables, representing concentrations of different molecular species across a biological probe. Apparently, this "comparative" approach accounts well for the biological mechanisms governing cellular protein networks. It suggests that cells organise the spatial structure of their protein networks in a highly non-random fashion, by tending to optimise their "mutual information content", and therefore most probably their efficiency.

We believe our method has a major advantage, in that it provides two different but closely intertwined kinds of information about the stacks of fluorescence images as follows.

(1) Biologically well-justified and fully user-independent threshold values obtained for each image in a stack can be very useful, as demonstrated in the case the MELK data used

in our study. Thus the threshold values obtained allow us to readily analyse the results from MELK technology, providing potentially important information about the spatial distribution of simultaneously present (or absent) combinations of molecular species within the biological object under investigation – information that can be used to support interpretation and visualisation of these MELK results.

(2) Our method measures statistical (positive and negative) associations between the random variables derived from a given image stack, or any subset of those variables, via the value of the *Multi-Information Function*.

Further developments could include an investigation of the statistical and geometric properties of the *Multi-Information Function*, and some improvement of the numerical optimisation procedure used to find the global maximum (cf. 2.2), to allow the processing of much larger amounts of data.

Acknowledgments

References

[1] A. Barysenka, *Copula-Based Methods of Exploring Statistical Dependence Between Fluorescent Markers*, PhD thesis, in preparation.

[2] Barysenka, A., *An Information Theoretic Thresholding Method for Detecting Protein Colocalizations in Stacks of Fluorescence Images*, J. Biotechnol. (2009), in press.

[3] M. Bode, M. Irmler, M. Friedenberger, C. May, K. Jung, C. Stephan, H. E. Meyer, C. Lach, R. Hillert, A. Krusche, J. Beckers, K. Marcus, W. Schubert, *Interlocking transcriptomics, proteomics and toponomics technologies for brain tissue analysis in murine hippocampus*, Proteomics, **8**:1170–8 (2008).

[4] T. Cover, *Elements Of Information Theory,* Wiley, 1991.

[5] A. Dress, T. Lokot, L.D. Pustyl'nikov, and W. Schubert, *Poisson Numbers and Poisson Distributions in Subset Surprisology*, Annals of Combinatorics **8**, 473 – 485 (2004).

[6] S. Kullback, *Information Theory and Statistics,* Dover, 1968.

[7] M. Friedenberger, M. Bode, A. Krusche, W. Schubert, *Fluorescence detection of protein clusters in individual cells and tissue sections by using toponome imaging system: sample preparation and measuring procedures*, Nat. Protoc., **2**:2285–94 (2007).

[8] R. Nelsen, *An Introduction to Copulas*, Springer, 1999.

[9] N. Otsu, *A threshold selection method from gray-level histograms*, IEEE Transactions on Systems, Man and Cybernetics, **9**:62–66 (1979).

[10] *Otsu's method*, http://en.wikipedia.org/wiki/Otsu's_method.

[11] *Kullback-Leibler divergence*, http://en.wikipedia.org/wiki/Kullback-Leibler_divergence.

[12] W. Schubert, *Topological Proteomics, Toponomics, MELK-Technology*, in *Proteomics of Microorganisms*, **83**, Springer 2003.

[13] W. Schubert, *Topological Proteomics Technology and Paradigm for Cell Invasion Dynamics*, J. Theor. Med., **4**:75–84 (2002).

[14] W. Schubert, B. Bonnekoh, A. J. Pommer, L. Philipsen, R. Böckelmann, Y. Malykh, H. Gollnick, M. Friedenberger, M. Bode, and A. W. M. Dress, *Analyzing proteome topology and function by automated multidimensional fluorescence microscopy*, Nat Biotechnol, **24**:1270–1278, 2006.

[15] W. Schubert, *Human toponome project*, Journal of Proteome Research, **7**:1806 (2008).

[16] W. Schubert, *The molecular face of prostate cancer*, Journal of Proteome Research, **8**:2616 (2009).

[17] W. Schubert, A. Gieseler, A. Krusche, R. Hillert, *Toponome mapping in prostate cancer: detection of 2000 cell surface protein clusters in a single tissue section and cell type specific annotation by using a three symbol code*, Journal of Proteome Research, **8**:2696–2707 (2009).

[18] W. Schubert, M. Friedenberger, M. Bode, A. Krusche, R. Hillert *Functional architecture of the cell nucleus: towards comprehensive toponome reference maps of apoptosis*,Biochim. Biophys. Acta, **1783**:2080–8 (2008).

[19] W. Schubert, M. Bode, R. Hillert, A. Krusche, M. Friedenberger *Toponomics and neurotoponomics: a new way to medical systems biology*, Expert Rev. Proteomics, **5**:361-9 (2008).

[20] W. Schubert, *Breaking the biological code*, Cytometry A, **71**:771–2 (2007).

[21] W. Schubert, *A three-symbol code for organized proteomes based on cyclical imaging of protein locations*, Cytometry A, **71**:352–60 (2007).

[22] M. Studeny, J. Vejnarova, *The multi-information function as a tool for measuring stochastic dependence*, Learning in Graphical Models, 261 (1998).

[23] M. Studeny, J. Vejnarova, *Information theoretical approach to constitution and reduction of medical data*, International Journal of Medical Informatics, **45**:65–74 (1997).