

Oscar Castillo, Janusz Kacprzyk, and
Witold Pedrycz (Eds.)

Soft Computing for Intelligent Control and Mobile Robotics

 Springer

Subjective Colocalization Analysis with Fuzzy Predicates

Pablo Rivas-Perea, Jose Gerardo Rosiles, and Wei Qian

The University of Texas El Paso,
Department of Electrical and Computer Engineering, El Paso TX 79968, USA

Abstract. Understanding the protein-to-protein interactions at the subcellular level, as well as other organic molecules, is crucial to explain cellular functions and to elucidate disease mechanisms. These interactions can be captured visually by overlapping the fluorescent microscopic images of two proteins tagged with fluorescent labeling agents that react to green and red wavelengths respectively. Interaction is determined by subjectively assessing the amount colocalization of green and red on the image composite based on the amount of yellow present in the image composite (i.e., green and red form yellow). Attempts to reduce the subjectivity of this process have focused on the computation of statistical coefficients and related methods. Even though statistical colocalization coefficients give a degree of correlation among the imaged proteins, they still need to be interpreted with subjective qualifiers like "high", "low", "strong", etc. Hence, there is no current agreement on the meaning of these coefficients among researchers. In this paper we propose the use of fuzzy linguistic variables to model the subjective interpretation of co-localization coefficients. Based on interpretations found in the literature, we produce a set of rules that map the coefficient values to a linguistic interpretation. The result of this work is a tool that provides an descriptive ensemble of coefficient interpretations that could guide researchers to a uniform interpretation colocalization criteria.

1 Introduction

Fluorescence microscopy makes possible the functional analysis of proteins and other other sub-cellular structures [1, 2, 3]. This imaging technique consists of tagging proteins of interest using fluorescent labeling agents which react to a specific wavelength. The spatial distribution of a protein over a cell is captured through an image sensor tuned to the fluorochrome wavelength.

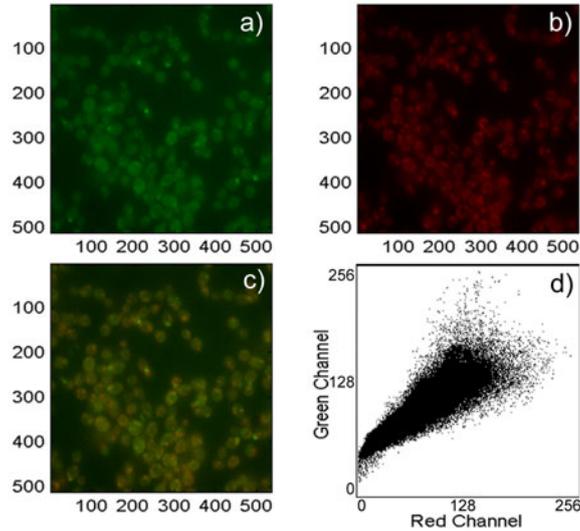


Fig. 1. An example image from the Yeast database. In (a) is shown the green channel. In (b) is shown the red channel, and in (c) the two joint channels. In (d) we have the scatter diagram showing the pixels distribution across the two channels.

Among the many uses of fluorescent microscopy images, it is possible to assess the interaction among two or more proteins at the sub-cellular level. Understanding these interactions is crucial to explain cellular functions and to elucidate disease mechanisms. The simultaneous presence of two proteins in the same sub-cellular compartment indicates a functional relationship for some specific biological process. Imaging the same specimen with two proteins tagged to express at different wavelengths (e.g., green and red), allows us to register the spatial distribution of the proteins simultaneously. By overlaying both images on the same plane, yellow regions would appear on those regions where red and green pixels spatially overlap. This overlaying method and the process of assessing the relative overlap (i.e., yellow regions) among the two proteins is known as *colocalization*.

An example using images from the yeast database [4, 5, 6] is shown in Figure 1, including the highlight of areas with strong visual colocalization. Biologists typically do a subjective assessment of colocalization through visual inspection of image overlays. Hence, colocalization results can vary from one person to another, or from one study to another. Attempts to reduce the subjectivity of this process have focused on the computation of statistical coefficients [7, 8, 9, 10] and related methods [11, 12, 13]. Even though statistical colocalization coefficients give a degree of correlation among the imaged proteins, they still need to be interpreted with subjective qualifiers like "high", "low", "strong", etc. Hence, there is no current agreement on the

meaning of these coefficients among researchers. In addition, they seem to be strongly influenced by the acquisition and post-processing steps applied to the images.

In this paper we propose the use of fuzzy linguistic variables [14, 15, 16, 17, 18] to model the subjective interpretation of co-localization coefficients. Based on interpretations found in the literature, we produce a set of rules that map the coefficient values to a linguistic interpretation. The result of this work is a tool that provides an descriptive ensemble of coefficient interpretations that could guide researchers to a uniform interpretation colocalization criteria. The proposed scheme can be described in two steps. First, using fluorescence microscopy images, a quantitative description of colocalization is obtained through the use of statistical coefficients reported in the literature [7, 8, 11, 12, 13, 19, 20] as features to fuzzy models. Second, based on the qualitative interpretation of different colocalization case studies, we constructed fuzzy linguistic variables to represent the quantitative results. This was followed by membership functions modeling. We tested our fuzzy model over the yeast and A431 cells image databases. The proposed model outputs natural language fuzzy predicates providing a consistent description of quantitative colocalization analysis results.

The paper is organized as follows. In Section 2 we give a brief introduction to the fluorescence microscopy imaging process. The quantitative colocalization feature extraction process is discussed in Section 3. In Section 4 we discuss the design our fuzzy logic-based subjective colocalization analysis model. Examples of the model utilization are presented in Section 5. We discuss our conclusions and subsequent work in Section 6.

2 An Overview of Fluorescence Microscopy

Fluorescence microscopy is an imaging technique where proteins (or some other molecule) are tagged using fluorescent labeling agents (fluorophores) which react to a specific wavelength. When different proteins on the same specimen are tagged, a multichannel image is generated, where each channel captures the contribution of each fluorophore. The image formation process can be approximated as a linear spatially-invariant system where the intensity distribution is modeled as [19]

$$I(n_1, n_2, n_3) \propto \int_{\mathbb{R}^3} \left| h_\lambda \left(\frac{n_1}{\hat{\zeta}} - u, \frac{n_2}{\hat{\zeta}} - v, \frac{n_3}{\hat{\zeta}} - w \right) \right|^2 \chi(u, v, w) du dv dw, \quad (1)$$

where $\lambda = \frac{hc}{E_{em}}$ is the emitted light wavelength, E_{em} is the energy level difference during light emission, h is Planck's constant, c is the speed of light, $\hat{\zeta}$ is the magnification of the microscope objective, χ is an object dependent function related to the light emission properties of the fluorophore.

The function $|h_\lambda|^2$ is known as the point spread function (PSF) of the microscope. It can be measured experimentally or modeled using the physical properties of the system. A common mathematical model is given by [19]

$$h_\lambda(n_1, n_2, n_3) = \int_{\mathbb{R}^2} P(u, v) e^{j2\pi n_3 \frac{u^2+v^2}{2\lambda\alpha^2}} e^{-j2\pi \frac{n_1 u + n_2 v}{\lambda\alpha}} dudv, \quad (2)$$

which represents the inverse Fourier transform of the circular aperture of the objective $P(u, v)$. The parameter α is the focal length of the objective and is inversely related to the numerical aperture (NA) of the microscope. A detailed discussion of this model can be found in [19].

As with other imaging techniques, images are acquired under non-ideal conditions that generate undesirable distortions or aberrations. In microscopy imaging, detection of light emissions is severely limited by the properties of photo sensors. Light detection is effectively down to counting the number of photons incident upon a photodetector. At such level, the photon emission is described by a Poisson random variable which generates noise contaminated images [19, 21, 22, 23]. Background shading is also a common distortion and can be caused by several factors: non uniform illumination, inhomogeneous detector sensitivity, dirt particles in the optics, nonspecific sample staining, and even auto-fluorescence. The background shading process can be modeled as

$$b(n_1, n_2) = I(n_1, n_2)a(n_1, n_2), \quad (3)$$

where $b(n_1, n_2)$ represents the product of the illumination $I(n_1, n_2)$ and the original sample $a(n_1, n_2)$ [19, 24]. Now, if we model the detector's gain and offset we have

$$c(n_1, n_2) = g(n_1, n_2)b(n_1, n_2) + o(n_1, n_2), \quad (4)$$

where $g(n_1, n_2)$ is some gain and $o(n_1, n_2)$ some offset. Using (3) to substitute in (4) we have

$$c(n_1, n_2) = g(n_1, n_2)I(n_1, n_2)a(n_1, n_2) + o(n_1, n_2). \quad (5)$$

These defects need to be corrected before analysis is performed. Digital enhancement and restoration techniques have been extensively studied [3, 9, 10, 25, 26]. In [27] we introduced a restoration method that recovers $a(n_1, n_2)$ given the noisy observation $c(n_1, n_2)$ described in (5). The proposed method is simple and performed better than traditional ones. Essentially the method consists of a local median filter (e.g. 3×3 followed by background subtraction. For an image consisting of red (R) and green (G) channels, a restored image $\hat{a}(n_1, n_2)$ is generated by

$$\hat{a}(n_1, n_2) = \begin{cases} \varrho_{[3 \times 3]} \{c(n_1, n_2)_G\} - \hat{o}(n_1, n_2)_G \\ \varrho_{[3 \times 3]} \{c(n_1, n_2)_R\} - \hat{o}(n_1, n_2)_R \end{cases}, \quad (6)$$

where $\varrho_{[3 \times 3]} \{\cdot\}$ is a 3×3 median filter is applied to each channel and $\hat{o}(n_1, n_2)$ is the offset estimate obtained as the average of a set of representative images from the database.

To assess restoration filter performance is computed using different metrics over synthetic and real life images. Consider the synthetic images shown

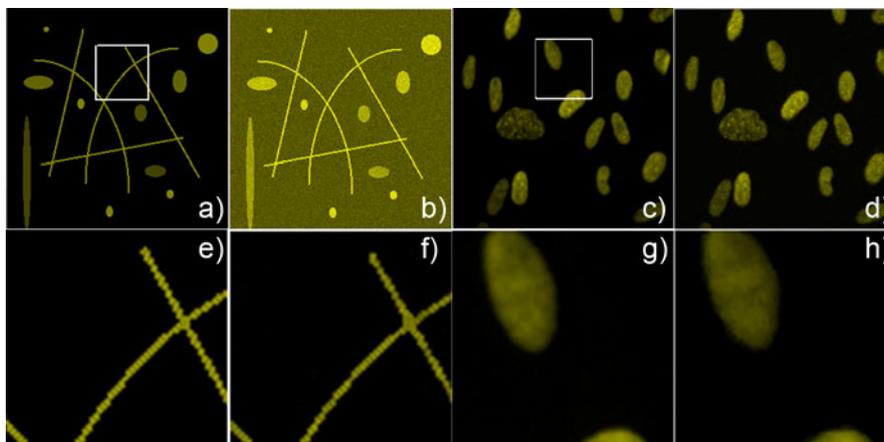


Fig. 2. In (a) is shown the “circles and lines” image, and its degraded version in (b). The “cells” image is shown in (c), and its degraded version in (d). In (e) is presented the original amplified section of circles and lines, and its restored image in (f). The original amplified cells in (g), and its restored image in (h). The results show very good performance over Poisson noise and offset, visually and quantitatively.

in Figure 2 (a) and (c). Synthetic images were degraded with a constant offset $o(n_1, n_2) = \eta$ added to each pixel. Then, images were contaminated with random Poisson noise. Figure 2 (a) is a typical case of study in image processing [9, 26] used to address the ability of a filter to recover edges, lines, and shape of the original image given a noisy observation as in (b). Likewise, Figure 2 (c) is representative of a microscopy imaging scenario, and the goal is to test the filter ability to recover texture, edges, and illumination given a degraded image as in (d). In the test of Figure 2 (b) we used $\eta = 89$ and in Figure 2 (d) $\eta = 11$. Clearly, the results over synthetic images shown in Figure 2 (f) and (h) are very good when compared to the true images in Figure 2 (e) and (g). Therefore, we proceeded to test restoration results over real life data.

For the case of real biological images, we evaluated the restoration filter with two databases. First, we used the yeast database [6] which contains 547 two channel image samples. It is a very good alternative to test the restoration methods for colocalization analysis. In Figure 3 (a) is shown an example of the content of the yeast database as well as its restored version using M33GM in Figure 3 (a).

The second database was reported in [9, 26] and consist of visible fluorescent fusion-proteins expressed in A431 cells and a fluorescent label. The erbB family of receptor tyrosine kinases includes the epidermal growth factor receptor (erbB1) and erbB3. These membrane proteins regulate cell growth and differentiation through binding of ligands to their extracellular domain,

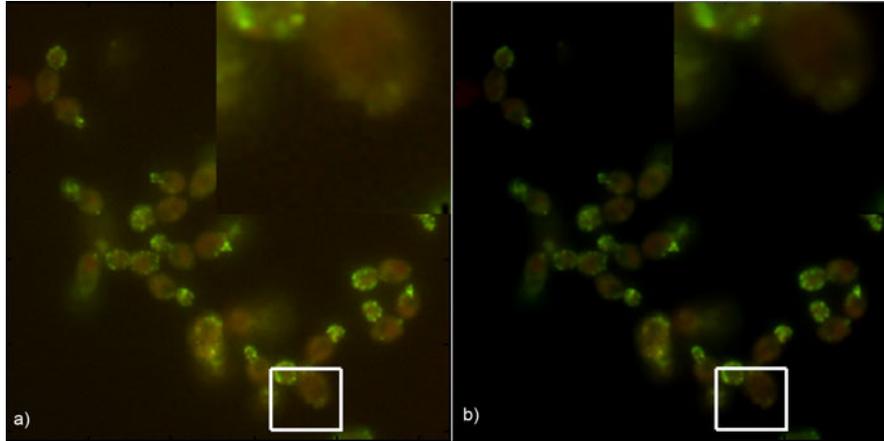


Fig. 3. In (a) is shown the noisy raw image from the Yeast database and in (b) its restored image

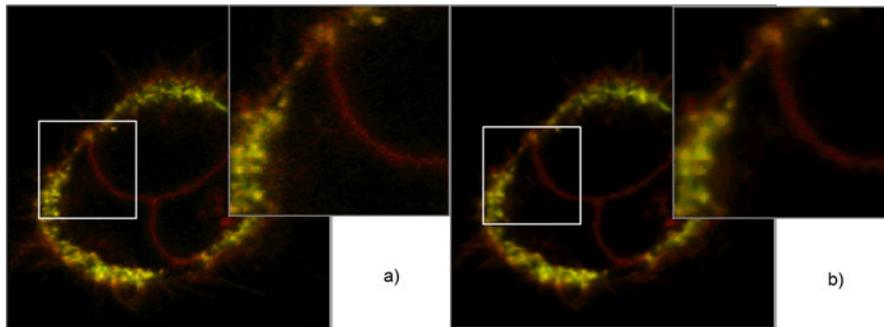


Fig. 4. Visible fluorescent fusion-proteins expressed in A431 cells. In (a) the noisy image, and in (b) the restored image

which activates the protein and initiates signalling. We identify this group of images as the A431 cells database. An example image is shown in Figure 4 (a), and the results of restoration are shown in Figure 4 (b). The filter effectively reduces Poisson noise, preserves the texture, and subtracts the offset.

3 Quantitative Colocalization Feature Extraction

The quantitative analysis of colocalization consists of computing the 2D signal spatial overlap across multiple channels. This high precision analysis allows researchers to understand the mechanisms of protein-to-protein

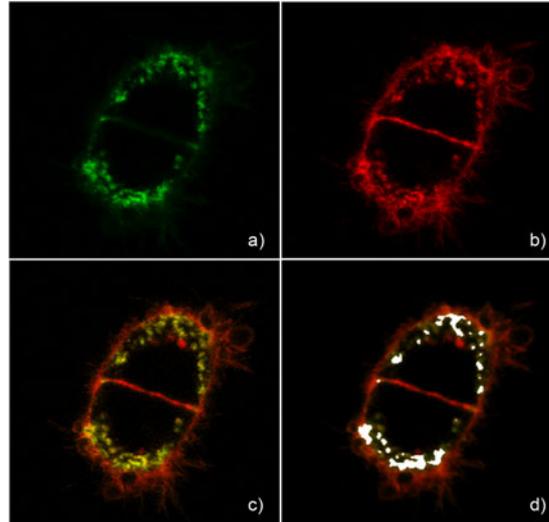


Fig. 5. An example of colocalization. In (a) is shown the green channel $G(n_1, n_2)$; in (b) is shown the red channel $R(n_1, n_2)$; in (c) are shown the superimposed channels; and in (d) are shown in white, the regions of more colocalization.

interactions. [3, 7, 8, 11, 12, 13, 19, 28]. As an example, consider the cellular image acquired by the fluorescence microscopy method; its green channel $G(n_1, n_2)$ is shown in Figure 5 (a), and its red channel $R(n_1, n_2)$ is shown in Figure 5 (b). We can study the cellular protein-to-protein interaction by observing and quantifying the superposition of Figure 5 (a) and Figure 5 (b) as in Figure 5 (c). The regions of more colocalization are highlighted in white on Figure 5 (d). Ultimately, visual assessment of colocalization is a subjective process highly dependent on the experience of the analyst. Attempts to reduce the subjectivity of this process have focused on the computation of statistical coefficients or parameters [7, 8, 10, 11, 25] and algorithms that exploit them [1, 19, 20, 28, 29]. There are specific parameters (i.e., features) used to estimate the degree of colocalization between two channels. We have identified 14 colocalization features in literature. In this paper, we will describe and use the five most popular features: Pearson's correlation coefficient, Overlap coefficient, Fraction of colocalizing regions, Mander's colocalization coefficient, and Intensity correlation coefficient.

Table 1 summarizes the properties of the coefficients that will be explained in detail next.

3.1 Pearson's Correlation Coefficient r_P

Also known as Correlation Coefficient, the Pearson's Correlation Coefficient is widely used and accepted, specially in regression applications. It provides

Table 1. Summary of features (coefficients) for quantitative colocalization

Name	Variables	Range	Ref.
Pearson's Corr. C.	r_P	$[-1, 1]$	[9, 10, 25]
Overlap C. (Mult. Meth.)	r	$[0, 1]$	[7, 12, 13]
Frac. Coloc. Reg. (O. C.)	k_1, k_2	vary	[7, 12, 13]
Mander's Coloc. C. - Gen.	M_1, M_2	$[0, 1]$	[7, 12, 13]
Intensity Corr. Quot.	ICQ	$[-0.5, 0.5]$	[7]

information about the relationship between the region of intensities and their distribution [7, 8, 9, 10, 11, 12, 13]. The Pearson's Correlation Coefficient r_P is defined as

$$r_P = \frac{\sum (R(n_1, n_2) - E[R]) (G(n_1, n_2) - E[G])}{\sqrt{\sum (R(n_1, n_2) - E[R])^2 \sum (G(n_1, n_2) - E[G])^2}}. \quad (7)$$

Its domain is between $[-1, 1]$. It will be -1 if the data has negative linear relationship, a particular meaning is hard to explain. It will be 0 (or close to zero) if the data has perfect exclusion or no linear relationship and the pixels are distributed along the entire scatter diagram. If the value is 1 it means perfect correlation; the data has a linear relationship. For the case shown in Figure 5 the value is $r_P = 0.7267$.

3.2 Overlap Coefficient r (The Multiply Method)

This coefficient dictates the overlap between two channels; it shows a degree of colocalization [12]. In contrast with Pearson's, this coefficient does not return negative values, does not average any pixel intensity, and it is not sensitive to intensity variations [11, 13]. However, the authors of [7] recommend its usage under the following condition: $\frac{\sum_{n_1, n_2} G(n_1, n_2)}{\sum_{n_1, n_2} R(n_1, n_2)} \approx 1$. The Overlap coefficient is formally defined as

$$r = \frac{\sum_{n_1, n_2} G(n_1, n_2) R(n_1, n_2)}{\sqrt{\sum_{n_1, n_2} G(n_1, n_2)^2 \sum_{n_1, n_2} R(n_1, n_2)^2}}, \quad (8)$$

where its domain is between $[0, 1]$. A value of 0 means that no pixels overlap; while 1 means that all the pixels overlap. The case when $r = 0.5$ implies that 50% of the pixels overlap with each other. For the case shown in Figure 5 $r = 0.8316$. It is clearly high because of the non-zero background content.

3.3 Fraction of Colocalizing Regions k_1, k_2 (Overlap Coefficients)

These coefficients represent the differences between each channel intensities [11, 12, 13]. This two coefficients overcome the problems generated from a restriction in the overlap coefficient in (8). However, k_1, k_2 are very sensitive to

the absolute fluorescent intensity. If one channel has been treated differently from the other, such that the total intensity vary (bleaching for instance), this will affect the coefficients [7]. We can denote the coefficients as follows

$$k_1 = \frac{\sum_{n_1, n_2} G(n_1, n_2)R(n_1, n_2)}{\sum_{n_1, n_2} G(n_1, n_2)^2}, \quad (9)$$

$$k_2 = \frac{\sum_{n_1, n_2} R(n_1, n_2)G(n_1, n_2)}{\sum_{n_1, n_2} R(n_1, n_2)^2}. \quad (10)$$

The individual range of values may vary. Results can be interpreted as some indicator of each antigen's contribution to colocalization areas.

3.4 Mander's Colocalization Coefficients M_1, M_2

The coefficients M_i , describe i th channel's contribution to colocalization using its intensity values. These coefficients are proportional to i th channel's fluorescence amount, relative to the i th channel total fluorescence. Such relationship is described in [7, 11, 12, 13] as

$$M_1 = \frac{\sum_{n_1, n_2} (G(n_1, n_2)|R(n_1, n_2) > 0)}{\sum_{n_1, n_2} G(n_1, n_2)}, \quad (11)$$

$$M_2 = \frac{\sum_{n_1, n_2} (R(n_1, n_2)|G(n_1, n_2) > 0)}{\sum_{n_1, n_2} R(n_1, n_2)}, \quad (12)$$

where M_1, M_2 are within the range $[0, 1]$. The meaning can be explained with the following example: if $M_1 = 1.0$ and $M_2 = 0.3$, it means that green channel pixels colocalize with red's, but only 30% of pixels in red channel colocalize with green's.

3.5 Intensity Correlation Quotient ICQ

Two images vary around their respective mean if their intensities vary in synchronous [7]. Therefore the *product of the differences from the mean* (PDM) will be positive for such images, $PDM > 0$. However, if the intensities vary asynchronously, the PDM will be negative, $PDM < 0$. The PDM analyzes the relationship between intensities, and is denoted as

$$PDM(n_1, n_2) = (G(n_1, n_2) - E[G]) (R(n_1, n_2) - E[R]), \quad (13)$$

where $E[G]$ and $E[R]$ denote the expected value for each channel respectively.

The ICQ value is based on the sign of the *PDM* [7]. The ICQ is defined as the PDM positive occurrences number ζ , divided by the negative occurrences ξ . Thus, the ICQ is the quotient denoted as

$$ICQ = \left(\frac{\zeta}{\xi} \right) - 0.5. \quad (14)$$

The range of *ICQ* falls between $[-0.5, 0.5]$. The coefficient value can be interpreted as follows: if the *ICQ* ≈ 0 , means random decoloration; if $-0.5 \leq ICQ < 0$, means segregated decoloration; and if $0 < ICQ \leq 0.5$ means a dependent decoloration. The latter case is an indicator of good colocalization.

4 Colocalization Subjective Quantification with Fuzzy Logic Theory

The colocalization study performer often would like to obtain a linguistic result from a coefficient rather than just a rough number. Consider Pearson's correlation coefficient $r_P = 0.865$ after some colocalization experiment. Then, the observer of the study would probably say, "Pearson's correlation coefficient is *high*." However, this is a subjective assessment that may not correspond to another observer whose definition of *high* may be a coefficient valued at 0.99. Hence, there is paucity of reference levels to map coefficient values to a human like quantification. The process of giving such subjective quantification is called *fuzzification* [18]. Prior knowledge regarding the variables to fuzzify is desired, but not mandatory. Information about the variable range and distribution is useful. In this section we address the problem of the subjective quantification by using fuzzy logic theory.

4.1 Linguistic Variables in Fuzzy Logic

A linguistic variable is a word represented by a fuzzy set [18]. For example, a linguistic variable can take an attribute like "*negative, positive, zero, high, low, more or less,*" etc. Each of this words has its own membership function.

Following the notation in [18], let us define a linguistic variable by the following quintuple

$$(x, T(x), U, G, M) \quad (15)$$

x : name of variable

$T(x)$: set of possible linguistic terms for x

U : set of universe of discourse

G : syntactic grammar that produces terms in $T(x)$

M : semantic rules which map terms in $T(x)$ to fuzzy sets in U .

Given the previous definitions, we introduce the translation of coefficients in Section 3 to linguistic variables. To distinguish between a crisp and a linguistic (fuzzy) variable we use the symbol $\hat{\cdot}$. For instance, the crisp Pearson’s correlation coefficient is denoted as r_P , while the linguistic Pearson’s correlation coefficient is defined as \hat{r}_P .

4.2 Fuzzy Pearson’s Correlation Coefficient

From [12] we know that r_P ’s universe of discourse is $[-1, 1]$. Now, we must find the possible words and their membership functions. In [12] it is found that -1 corresponds to “negative” correlation; so, let $\mu_{neg}^{\hat{r}_P}(u)$ denote this possible term for \hat{r}_P . From [7] and [12] we also can find $\mu_{lit}^{\hat{r}_P}(u)$ as a “little” correlation, with an average value close to 0; as well as “reasonable” $\mu_{rea}^{\hat{r}_P}(u)$, with an average value close to 0.346. Finally, we found “strong” correlation $\mu_{str}^{\hat{r}_P}(u)$, centered at 0.920. In [26] can be found another two possible terms. The first is “high” $\mu_{hig}^{\hat{r}_P}(u)$, in the range $[0.685, 0.877]$, with an average value of 0.761. The second term is “low” $\mu_{low}^{\hat{r}_P}(u)$, in the range $[0.434, 0.615]$ with average 0.516.

At universe of discourse upper and lower limits, “s membership functions” (*SMF*) and “z membership functions” (*ZMF*) are commonly used since they consider the inclusion of such limits. In comparison, a Gaussian membership function does not fully cover the limits of the universe of discourse. Formally a “z membership function” is defined as

$$\mu(u) = \begin{cases} 1, & u \leq a \\ 1 - 2 \left(\frac{u-a}{b-a} \right)^2, & a \leq u \leq \frac{a+b}{2} \\ 2 \left(b - \frac{u}{b-a} \right)^2, & \frac{a+b}{2} \leq u \leq b \\ 0, & u \geq b \end{cases} \tag{16}$$

where parameters a and b define the start and end of the function respectively.

Similarly, an “s membership function” is formally defined as

$$\mu(u) = \begin{cases} 0, & u \leq a \\ 2 \left(\frac{u-a}{b-a} \right)^2, & a \leq u \leq \frac{a+b}{2} \\ 1 - 2 \left(b - \frac{u}{b-a} \right)^2, & \frac{a+b}{2} \leq u \leq b \\ 1, & u \geq b \end{cases} \tag{17}$$

where the parameters a and b define the start and end of the function respectively. A Gaussian-shaped membership function (*GMF*) falls in the family of the exponential and radial basis functions; and it is defined as

$$\mu(u) = e^{-\frac{(u-c)^2}{2\sigma^2}}, \tag{18}$$

where c is the centering parameter, and σ is the spread parameter. A similar function is the π -shaped membership function (*PIMF*) defined as

$$\mu(u) = \frac{1}{1 + \left| \frac{u-c}{a} \right|^{2b}}, \tag{19}$$

with parameters a and b defining the start and end of the function respectively; and c as the centering parameter. Now, following the definition in (15), we can define \hat{r}_P as follows

$$\begin{aligned} \hat{r}_P = & \tag{20} \\ & x : \text{Pearson's Correlation} \\ T(x) : & \{\text{negative, little, reasonable, strong, high, low}\} \\ U : & [-1, 1] \\ G(x) : & T^{i+1} = \{\text{strong}\} \vee \{\text{very } T^i\} \\ M : & \left\{ \begin{array}{l} (u, \mu_{neg}^{\hat{r}_P}(u)), (u, \mu_{lit}^{\hat{r}_P}(u)), \\ (u, \mu_{rea}^{\hat{r}_P}(u)), (u, \mu_{str}^{\hat{r}_P}(u)), \\ (u, \mu_{high}^{\hat{r}_P}(u)), (u, \mu_{pos}^{\hat{r}_P}(u)), \\ (u, \mu_{low}^{\hat{r}_P}(u)) \mid u \in U \end{array} \right\} \end{aligned}$$

where the operator \vee denotes the fuzzy operator max; $\mu_{neg}^{\hat{r}_P}(u)$ uses a function of the type defined in (16) with parameters $a = -1$, and $b = 0$; $\mu_{lit}^{\hat{r}_P}(u)$ uses a function of the type defined in (18) with parameters $c = 0$, and $\sigma = 0.2$; $\mu_{rea}^{\hat{r}_P}(u)$ uses a function of the type defined in (18) with parameters $c = 0.346$, $\sigma = 0.2$; $\mu_{str}^{\hat{r}_P}(u)$ uses a function of the type defined in (18) with parameters $c = 0.92$, $\sigma = 0.2$; $\mu_{high}^{\hat{r}_P}(u)$ uses a function of the type defined in (18) with parameters $c = 0.761$, $\sigma = 0.2$; $\mu_{low}^{\hat{r}_P}(u)$ uses a function of the type defined in (18) with parameters $c = 0.516$, $\sigma = 0.2$; and finally $\mu_{pos}^{\hat{r}_P}(u)$ uses a function of the type defined in (17) with parameters $a = 0.434$, and $b = 1$.

4.3 Fuzzy Overlap Coefficient

In [7, 12] we can find that the universe of discourse for the linguistic variable \hat{r} is $[0, 1]$. From [7] we obtain $\mu_{ran}^{\hat{r}}(u)$ as a “*random colocalization*” overlap, centered near to 0.5; as well as “*high colocalization*” $\mu_{high}^{\hat{r}}(u)$, when it is 1. Finally, we found “*low colocalization*” $\mu_{low}^{\hat{r}}(u)$, when the value reaches 0. So, following the notation in (15), we can define \hat{r} as

$$\begin{aligned}
 \widehat{r} &= & (21) \\
 x &: \text{Overlap Coefficient} \\
 T(x) &: \left\{ \begin{array}{l} \text{random colocalization,} \\ \text{high colocalization,} \\ \text{low colocalization} \end{array} \right\} \\
 U &: [0, 1] \\
 G(x) &: T^{i+1} = \{\text{low}\} \vee \{\text{very } T^i\} \\
 M &: \left\{ \begin{array}{l} (u, \mu_{ran}^{\widehat{r}}(u)), \\ (u, \mu_{hig}^{\widehat{r}}(u)), \\ (u, \mu_{low}^{\widehat{r}}(u)) \mid u \in U \end{array} \right\}
 \end{aligned}$$

where $\mu_{ran}^{\widehat{r}}(u)$ uses (19) for $a = 0.25$, $b = 3$, and $c = 0.5$; $\mu_{hig}^{\widehat{r}}(u)$ uses (17) for $a = 0.5$, and $b = 1$; and finally, $\mu_{low}^{\widehat{r}}(u)$ uses (16) for $a = 0$, and $b = 0.5$.

4.4 Fuzzy Fraction of Colocalizing Regions (Overlap Coefficients)

For this coefficient, the most used is the k_1 coefficient since it shows the ratio between channel red versus green [7]. As specified previously this variable has a variable universe of discourse; however, the most common values lie close to one. From this finding we can define the linguistic variable \widehat{k}_1 over the range $[-\infty, \infty]$. The variable \widehat{k}_1 has one membership function centered at one and is denoted as $\mu_{val}^{\widehat{k}_1}(u)$ which is a “valid” ratio between channels. Following the notation in (15), we can define \widehat{k}_1 as

$$\begin{aligned}
 \widehat{k}_1 &= & (22) \\
 x &: \text{Ch.Red/Ch.Green ratio} \\
 T(x) &: \{\text{valid, not valid}\} \\
 U &: [-\infty, \infty] \\
 G(x) &: \overline{T}(x) = 1 - T(x) \\
 M &: \left\{ \begin{array}{l} (u, \mu_{val}^{\widehat{k}_1}(u)), \\ (u, \mu_{nov}^{\widehat{k}_1}(u)) \mid u \in U \end{array} \right\}
 \end{aligned}$$

where $\mu_{val}^{\widehat{k}_1}(u)$ uses (18), for $c = 1$, and $\sigma = \frac{1}{3}$; while $\mu_{nov}^{\widehat{k}_1}(u)$ is just $1 - \mu_{val}^{\widehat{k}_1}(u)$.

4.5 Fuzzy Mander’s Colocalization Coefficients - General

The colocalization coefficients M_1, M_2 from Mander’s are defined over the universe of discourse $[0, 1]$ as explained previously, thus the linguistic variables $\widehat{M}_1, \widehat{M}_2$ will be defined in this universe as well. In [7] is described the word “significant” $\mu_{sig}^{\widehat{M}_1}(u)$ for a value of 0.746. Also we can find the word “not much” $\mu_{nom}^{\widehat{M}_1}(u)$ for 0.155, as well as “high” $\mu_{hig}^{\widehat{M}_1}(u)$ for the range

of $[0.879, 0.993]$ (with mean value 0.936). Similarly, we can find the word “strong” $\mu_{str}^{\widehat{M}_1}(u)$ defined over the range $[0.970, 0.998]$ (with mean value 0.984). We also find $\mu_{noc}^{\widehat{M}_1}(u)$ which is “not coincident” correlation for values less than 0.008. Finally the word “hard” $\mu_{har}^{\widehat{M}_1}(u)$ is found for a value of 0.590. Following the notation in (15), we define \widehat{M}_1 as

$$\begin{aligned} \widehat{M}_1 = & \tag{23} \\ & x : \text{Colocalization Coefficient for Red Channel} \\ T(x) : & \left\{ \begin{array}{l} \text{not coincident, not much, hard,} \\ \text{significant, high, strong} \end{array} \right\} \\ U : & [0, 1] \\ G(x) : & T^{i+1} = \{\text{strong}\} \vee \{\text{very } T^i\} \end{aligned}$$

$$M : \left\{ \begin{array}{l} \left(u, \mu_{noc}^{\widehat{M}_1}(u) \right), \\ \left(u, \mu_{nom}^{\widehat{M}_1}(u) \right), \left(u, \mu_{har}^{\widehat{M}_1}(u) \right), \\ \left(u, \mu_{sig}^{\widehat{M}_1}(u) \right), \left(u, \mu_{hig}^{\widehat{M}_1}(u) \right), \\ \left(u, \mu_{str}^{\widehat{M}_1}(u) \right) \mid u \in U \end{array} \right\}$$

where $\mu_{noc}^{\widehat{M}_1}(u)$ uses (16) for $a = 0.008$ and $b = 0.155$; $\mu_{nom}^{\widehat{M}_1}(u)$ uses (18) for $c = 0.155$ and $\sigma = 0.2$; $\mu_{har}^{\widehat{M}_1}(u)$ uses (18) for $c = 0.590$ and $\sigma = 0.2$; $\mu_{sig}^{\widehat{M}_1}(u)$ uses (18) for $c = 0.746$ and $\sigma = 0.2$; $\mu_{hig}^{\widehat{M}_1}(u)$ uses (18) for $c = 0.936$ and $\sigma = 0.2$; finally $\mu_{str}^{\widehat{M}_1}(u)$ uses (17) for $a = 0.746$ and $b = 0.984$. The linguistic variable \widehat{M}_2 is defined in the same way as \widehat{M}_1 , but the only difference is the value for x which is “Colocalization Coefficient for Green Channel.”

4.6 Fuzzy Intensity Correlation Quotient Coefficient

Since the *ICQ* coefficient [7, 30] is defined over the universe of discourse $[-0.5, 0.5]$, the linguistic variable \widehat{ICQ} will be defined in this universe as well. In [7] is described the word “segregated” $\mu_{seg}^{\widehat{ICQ}}(u)$ for a values in the range $[-0.5, 0]$. It is found as well the word “random” $\mu_{ran}^{\widehat{ICQ}}(u)$ for values near to zero. Finally the word “dependent” $\mu_{dep}^{\widehat{ICQ}}(u)$ is found for values on the range $[0, 0.5]$. Following the notation in (15), we define \widehat{ICQ} as

$$\begin{aligned}
\widehat{ICQ} = & \quad (24) \\
x : & \text{Intensity Correlation Quotient} \\
T(x) : & \{\text{segregated, random, dependent}\} \\
U : & [-0.5, 0.5] \\
G(x) : & T^{i+1} = \{\text{segregated}\} \vee \{\text{very } T^i\} \\
M : & \left\{ \begin{array}{l} \left(u, \mu_{seg}^{\widehat{ICQ}}(u) \right), \left(u, \mu_{ran}^{\widehat{ICQ}}(u) \right), \\ \left(u, \mu_{dep}^{\widehat{ICQ}}(u) \right) \Big| u \in U \end{array} \right\}
\end{aligned}$$

where $\mu_{seg}^{\widehat{ICQ}}(u)$ uses (16) for $a = -0.5$ and $b = 0.5$; $\mu_{ran}^{\widehat{ICQ}}(u)$ uses (18) for $c = 0$ and $\sigma = \frac{1}{8}$; and finally $\mu_{dep}^{\widehat{ICQ}}(u)$ uses (17) for $a = -0.5$ and $b = 0.5$.

4.7 Fuzzy Predicates from Linguistic Variables

Fuzzy predicates are a semantic representation of a sentence. For instance, “It is cold outside.” Here we identify “cold” as the linguistic variable. In colocalization analysis, we defined fuzzy predicates. As an example, consider the fuzzy variables and the fuzzy predicates shown in Table 2. The syntax is flexible so that the colocalization study performer can pick the combination that best represent the human-like reasoning or help the best in human-like decision making. Human-like knowledge representation can be achieved using ordinary fuzzy predicates. Such fuzzy predicates depend on linguistic variables. For the particular case of quantitative colocalization analysis through fluorescence microscopy, fuzzy membership functions were created according to the cases found in literature. In Figure 6 we present a summary of the modeled membership functions. The relevance of this research can be summarized in two parts: first, we addressed the inaccurate subjective human interpretation of colocalization, providing a formal review of quantitative colocalization coefficients; second, using fuzzy logic, we designed linguistic variables

Table 2. Definition of linguistic variables and fuzzy predicate examples

Fuzzy Predicate Definition	Fuzzy Predicate Example
“ $\widehat{r}_P[x]$ is $\widehat{r}_P[T(x)]$ ”	Pearson’s Correlation is strong
“ $\widehat{r}[x]$ is showing $\widehat{r}[T(x)]$ ”	“Overlap Coefficient is showing high colocalization”
“ $\widehat{k}_1[x]$ is $\widehat{k}_1[T(x)]$ ”	“Ch.Red/Ch.Green ratio is not valid”
“ $\widehat{M}_1[x]$ shows $\widehat{M}_1[T(x)]$ colocalization”	“Colocalization Coefficient for Red Channel is showing significant colocalization”
“ $\widehat{ICQ}[x]$ shows $\widehat{ICQ}[T(x)]$ staining”	“Intensity Correlation Quotient Coefficient shows random staining”

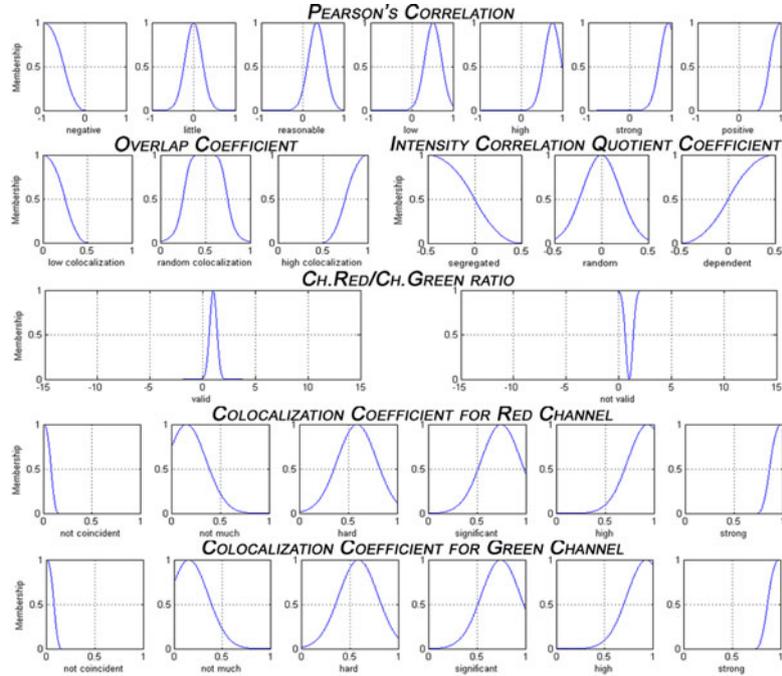


Fig. 6. Graphical representation of the linguistic variables, and its associated membership functions modeled from a literature review

based on literature review and provided an accurate human-like colocalization quantification.

4.8 Experimental Results and Discussion

We performed several experiments in different databases such as Yeast and A431 cells. The experiments consist of computing the previously described coefficients and using them as inputs to a fuzzy system. The fuzzy system consists of five linguistic variables that take its associated coefficient numerical value and produce a set of human like colocalization quantification. The linguistic variables x are constructed in the form of (15). The output of each linguistic variable is computed by taking the maximum valued membership function associated to that linguistic variable. Then, we use the linguistic term $T(x)$ associated to the maximum valued membership function to construct a human-like sentence (fuzzy predicate). This procedure can be formalized in four steps:

Step 1. Image restoration. Channels $R(n_1, n_2)$ and $G(n_1, n_2)$ are restored using the filter discussed in Section 2 to obtain $\hat{c}(n_1, n_2)$.

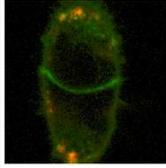
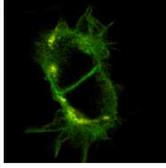
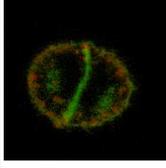
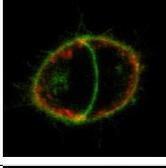
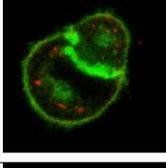
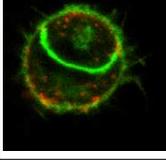
Table 3. Fuzzy predicates obtained from samples in Figures 1, 3, 4, and 5

Figure #	Fuzzy Predicates.
Figure 1	Pearson's Correlation is strong. Overlap Coefficient is showing high colocalization. Ch.Red/Ch.Green ratio is not valid. Colocalization Coefficient for Red Channel shows strong colocalization. Colocalization Coefficient for Green Channel shows strong colocalization. Intensity Correlation Quotient Coefficient shows segregated staining.
Figure 3	Pearson's Correlation is strong. Overlap Coefficient is showing high colocalization. Ch.Red/Ch.Green ratio is not valid. Colocalization Coefficient for Red Channel shows strong colocalization. Colocalization Coefficient for Green Channel shows strong colocalization. Intensity Correlation Quotient Coefficient shows segregated staining.
Figure 4.	Pearson's Correlation is strong. Overlap Coefficient is showing high colocalization. Ch.Red/Ch.Green ratio is not valid. Colocalization Coefficient for Red Channel shows strong colocalization. Colocalization Coefficient for Green Channel shows strong colocalization. Intensity Correlation Quotient Coefficient shows segregated staining.
Figure 5.	Pearson's Correlation is high. Overlap Coefficient is showing high colocalization. Ch.Red/Ch.Green ratio is not valid. Colocalization Coefficient for Red Channel shows strong colocalization. Colocalization Coefficient for Green Channel shows strong colocalization. Intensity Correlation Quotient Coefficient shows segregated staining.

Step 2. Feature extraction. Features proposed in Section 3 are extracted from $\widehat{c}(n_1, n_2)$ to form a row vector $F = [r_P, r, [k_1, k_2], [M_1, M_2], ICQ]^T$.

Step 3. Fuzzification. Features vector F is fuzzified according to the linguistic variable associated to each feature. Follows to construct a row vector $\widehat{F} = [\widehat{r}_P, \widehat{r}, [\widehat{k}_1, \widehat{k}_2], [\widehat{M}_1, \widehat{M}_2], \widehat{ICQ}]^T$.

Table 4. Fuzzy predicates obtained from the A341 cells database

	<p>Pearson's Correlation is high. Overlap Coefficient is showing high colocalization. Ch.Red/Ch.Green ratio is not valid. Colocalization Coefficient for Red Channel shows strong colocalization. Colocalization Coefficient for Green Channel shows strong colocalization. Intensity Correlation Quotient Coefficient shows segregated staining.</p>
	<p>Pearson's Correlation is strong. Overlap Coefficient is showing high colocalization. Ch.Red/Ch.Green ratio is not valid. Colocalization Coefficient for Red Channel shows strong colocalization. Colocalization Coefficient for Green Channel shows strong colocalization. Intensity Correlation Quotient Coefficient shows segregated staining.</p>
	<p>Pearson's Correlation is high. Overlap Coefficient is showing high colocalization. Ch.Red/Ch.Green ratio is not valid. Colocalization Coefficient for Red Channel shows strong colocalization. Colocalization Coefficient for Green Channel shows strong colocalization. Intensity Correlation Quotient Coefficient shows segregated staining.</p>
	<p>Pearson's Correlation is high. Overlap Coefficient is showing high colocalization. Ch.Red/Ch.Green ratio is not valid. Colocalization Coefficient for Red Channel shows strong colocalization. Colocalization Coefficient for Green Channel shows strong colocalization. Intensity Correlation Quotient Coefficient shows segregated staining.</p>
	<p>Pearson's Correlation is low. Overlap Coefficient is showing random colocalization. Ch.Red/Ch.Green ratio is not valid. Colocalization Coefficient for Red Channel shows strong colocalization. Colocalization Coefficient for Green Channel shows strong colocalization. Intensity Correlation Quotient Coefficient shows segregated staining.</p>
	<p>Pearson's Correlation is low. Overlap Coefficient is showing random colocalization. Ch.Red/Ch.Green ratio is not valid. Colocalization Coefficient for Red Channel shows strong colocalization. Colocalization Coefficient for Green Channel shows strong colocalization. Intensity Correlation Quotient Coefficient shows segregated staining.</p>

Step 4. Fuzzy predicates. For each element of the fuzzyfied features vector \hat{F} the linguistic term $T(x)$ is extracted according to the semantic rules M associated to each linguistic variable. Then, fuzzy predicates are constructed according to the skeleton shown in Table 2.

Let us consider the case of the samples shown in Figures 1, 3, 4, and 5. The fuzzy predicates obtained from these cases are presented in Table 3.

Clearly, fuzzy predicates alleviate the need for quantitative results in natural language. Consider the experiments with the A341 cells database shown in Table 4. These cases are also successful experiments in the construction of colocalization fuzzy predicates.

The results shown represent a paradigm change in the quantitative colocalization analysis since current problems are solved visually and by human analysis. In this paper is proposed the usage of features and fuzzy linguistic variables to provide an invariant human-independent result of colocalization. Results provide a descriptive set of colocalization coefficient interpretations that could lead researchers to a uniform and consistent interpretation colocalization.

5 Conclusion

In this document we addressed the subjectivity problem in quantitative colocalization analysis through fluorescence microscopy. We presented the most common quantitative colocalization coefficients reported in literature. These coefficients are utilized to model a fuzzy system. Linguistic variables are modeled using fuzzy logic theory from the reported colocalization coefficients. Fuzzy membership functions were constructed based on literature review. The proposed fuzzy model provides natural language quantitative colocalization results through the evaluation of membership functions associated with linguistic variables. Our model demonstrated successful assessment of quantitative colocalization with natural language results. Results over the Yeast and A431 cells database show consistency in the computations and in the quantitative colocalization assessment. The proposed model can be utilized in the field of biological sciences where confocal fluorescence microscopy techniques are used in the analysis of protein-to-protein interactions. This will alleviate the paucity of formal quantitative colocalization analysis using novel computational intelligence methods. The proposed model provides a human-like ensemble of colocalization coefficient interpretations that could lead researchers to a uniform and consistent interpretation colocalization.

Acknowledgement

The authors want to thank to the National Council of Science and Technology, CONACyT Mexico, for partially supporting this research under grant 193324. The authors acknowledge Dr. Filip Rooms for providing the A431 cell images.

References

1. Savio, E., Goldhaber, J.I., Bridge, J.H.B., Sachse, F.B.: A framework for analyzing confocal images of transversal tubules in cardiomyocytes. In: Sachse, F.B., Seemann, G. (eds.) FIHM 2007. LNCS, vol. 4466, pp. 110–119. Springer, Heidelberg (2007)

2. Matula, P., Kozubek, M., Matula, P.: Applications of image registration in human genome research. In: ECCV Workshops CVAMIA and MMBIA, pp. 376–384 (2004)
3. Lukas Landmann, P.M.: Colocalization analysis yields superior results after image restoration. *Microsc. Res. Tech.* 64, 103–112 (2004)
4. Huh, W., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., O’Shea, E.K.: Global analysis of protein localization in budding yeast. *Nature* 425, 686–691 (2003)
5. Ghaemmaghami, S., Huh, W., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O’Shea, E.K., Weissman, J.S.: Global analysis of protein expression in yeast. *Nature* 425, 737–741 (2003)
6. Howson, R., Huh, W.K., Ghaemmaghami, S., Falvo, J.V., Bower, K., Belle, A., Dephoure, N., Wykoff, D.D., Weissman, J.S., O’Shea, E.K.: Construction, verification and experimental use of two epitope-tagged collections of budding yeast strains. *Comp. Funct. Genomics* 6, 2–16 (2005)
7. McMaster Biophotonics Facility (30/11/2006). Colocalization, p. 20 (November 1, 2008)
http://www.macbiophotonics.ca/PDF/MBF_colocalisation.pdf
8. Indiana Center for Biological Microscopy (10/05/2007); Analysis of colocalization using metamorph, p. 7 (November/01, 2008),
<http://www.nephrology.iupui.edu/imaging/tutorials/AnalysisofColocalizationusingMetamorph.pdf>
9. Rooms, F.: Nonlinear Techniques in Image Restoration applied to Confocal Microscopy. In: Faculty of Engineering, Department TELIN (June 10, 2005)
10. Adler, J., Parmryd, I.: Letter to the Editor. *Microsc.* 227, 83 (2007)
11. MediaCybernetics (22/03/2002), Application note 1: Colocalization of fluorescence probes, pp. 20 (November/01, 2008),
<http://www.spectraservices.com/Merchant2/pdf/AppInfo-CoLocFluorProbes.pdf>
12. Zinchuk, V., Zinchuk, O., Okada, T.: Quantitative colocalization analysis of multicolor confocal immunofluorescence microscopy images: pushing pixels to explore biological phenomena. *Acta Histochem. Cytochem.* 40, 101–111 (2007)
13. CoLocalization Research Software (02/11/2008), CoLocalizer pro user guide. ver. 2.5., pp. 67 (November/01, 2008), <http://homepage.mac.com/colocalizerpro/Resources/CoLocalizerProUserGuide.pdf>
14. Ding, Y.: A New Obfuscation Scheme in Constructing Fuzzy Predicates. In: WRI World Congress on Software Engineering, WCSE 2009, May 19–21, vol. 4, pp. 379–382 (2009)
15. Drobics, M., Adlassnig, K.-P.: Extending the Medical Concept of Reference Intervals Using Fuzzy Predicates. In: International Conference on Computational Intelligence for Modelling Control and Automation, December 10–12, pp. 603–608 (2008)
16. Prade, H., Serrurier, M.: Getting adaptability or expressivity in inductive logic programming by using fuzzy predicates. In: Proceedings of 2004 IEEE International Conference on Fuzzy Systems, July 25–29, vol. 1, pp. 73–77 (2004)
17. Bosc, P., Pivert, O.: An approach for a hierarchical aggregation of fuzzy predicates. In: Second IEEE International Conference on Fuzzy Systems, vol. 2, pp. 1231–1236 (1993)
18. Tsoukalas, L.H., Uhrig, R.E.: *Fuzzy and Neural Approaches in Engineering*, p. 587. Wiley, New York (1997)

19. Wu, Q., Merchant, F., Castleman, K.R.: *Microscope Image Processing*, p. 548. Elsevier/Academic Press, Amsterdam (2008)
20. Agnati, L.F., Fuxe, K., Torvinen, M., Genedani, S., Franco, R., Watson, S., Nussdorfer, G.G., Leo, G., Guidolin, D.: New Methods to Evaluate Colocalization of Fluorophores in Immunocytochemical Preparations as Exemplified by a Study on A2A and D2 Receptors in Chinese Hamster Ovary Cells. *J. Histochem. Cytochem.* 53, 941–953 (2005)
21. Stark, H., Woods, J.W.: *Probability and Random Processes with Applications to Signal Processing*, 3rd edn. Prentice Hall, Englewood Cliffs (2001)
22. Krishnan, V.: *Probability and Random Processes*. John Wiley, Chichester (2006)
23. Gubner, J.A.: *Probability and Random Processes for Electrical and Computer Engineers*. Cambridge University Press, Cambridge (2006)
24. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 2nd edn. Prentice Hall, Englewood Cliffs (2002)
25. Adler, J., Bergholm, F., Pagakis, S., Parmryd, I.: Noise and Colocalization in Fluorescence Microscopy: Solving a Problem. *Microsc. and Analysis* 22 (2008)
26. Rooms, F., Philips, W., Lidke, D.S.: Simultaneous degradation estimation and restoration of confocal images and performance evaluation by colocalization analysis. *J. Microscopy* 218(1), 22–36 (2005)
27. Rivas-Perea, P., Rosiles, J.G., Qian, W.: Self organizing maps for class discovery in the quantitative colocalization analysis feature space. In: *Proc. of IEEE International Joint Conference on Neural Networks*, pp. 2678–2684 (2009)
28. Bolte, S., Cordelieres, F.P.: A guided tour into subcellular colocalization analysis in light microscopy. *Journal of Microscopy* 224(3), 213–232
29. Costes, S.V., Daelemans, D., Cho, E.H., Dobbin, Z., Pavlakis, G., Lockett, S.: Automatic and Quantitative Measurement of Protein-Protein Colocalization in Live Cells. *Journal of Biophysics* 86(6), 3993–4003
30. Wu, Y., Eghbali, M., Ou, J., Li, M., Toro, L., Stefani, E.: Quantitative Determination of Spatial Protein-protein Proximity in Fluorescence Confocal Microscopy. *Biological Physics* (2009)