Depth of Valleys Accumulation Algorithm for Object Detection

Jorge Bernal, Javier Sánchez and Fernando Vilariño
Computer Vision Center, Campus Universitat Autònoma de Barcelona, Edifici O, 08193 Bellaterra, Barcelona, Spain

Abstract. This work aims at detecting in which regions the objects in the image are by using information about the intensity of valleys, which appear to surround objects in images where the source of light is in the line of direction than the camera. We present our depth of valleys accumulation method, which consists of two stages: first, the definition of the depth of valleys image which combines the output of a ridges and valleys detector with the morphological gradient to measure how deep is a point inside a valley and second, an algorithm that denotes points of the image as interior to objects those which are inside complete or incomplete boundaries in the depth of valleys image. To evaluate the performance of our method we have tested it on several application domains. Our results on object region identification are promising, specially in the field of polyp detection in colonoscopy videos, and we also show its applicability in different areas.

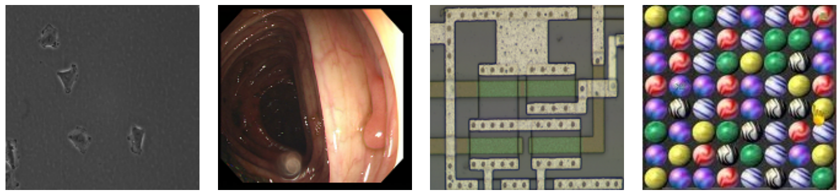
Keywords. Object Recognition, Object Region Identification, Image Analysis, Image Processing

1. Introduction

Object recognition is a recurring challenge in computer vision and artificial intelligence, concerned on answering questions such as what objects are in the picture and where? Currently it remains challenging mainly because the high variations that real-world images present. We could see object recognition schema as a three parts scheme: object detection, object description and object classification. Object detection methods find out where objects are in the image, no matter what they are. Object recognition algorithms must overcome some of the problems such as partial occlusions, viewpoint changes or variation in the illumination. Some of the available solutions in the context of pattern recognition need of the use of an image segmentation procedure followed by a machine learning approach to perform the final classification.

In this paper we present a novel object detection method that, without any image segmentation, detects which region of the image is more likely to contain an object inside by using valleys information. Valleys surround objects in images where the camera and the light source are in the same (or opposite) direction, which happens in microscopic, X-ray or endoscopic images, just to name a few (see Figure 1).

Our first contribution extends the definition of the depth of valleys image [3], which uses as an approximation that the boundaries that surround objects in image can be seen as intensity valleys. The problem with those intensity valleys is that in many cases they



(a) (b) (c) (d)

Figure 1. Example images (a) cells (b) endoscopic image (c) integrated circuit board (d) marbles.

are not closed and, therefore, we can not know if they are defining the boundary of an object or they are just non-important boundaries. Our method uses an approximation denoted as the depth of valleys image that not only uses the information about where valleys are but also their intensity so the points that belong to the boundaries of objects have a high value in the depth of valley image and the points interior to objects have a lower value in the depth of valleys image.

Our second contribution, ellipse fitting depth of valleys accumulation (EF-DOVA) consists of finding which points on the image belong to objects and which do not, by checking if they are surrounded, in a high number of directions by points with high value in the depth of valleys image.

The results of our method can be used for several applications that need of an object detection, as a way to indicate which parts or region of the image are more likely to contain objects. In this paper we show how our method can be applied in several real-world scenarios such as cell detection and polyp detection in colonoscopy videos. Both contributions are meant to work together in order to build a complete object detection algorithm but the accumulation algorithm, as we will show for cell detection, could also be applied to images where a edges or blob detection is more suitable (Figure 1 (d)).

The structure of the paper is as it follows: in Section 2 we present the ideas extracted from other papers that inspired our work. In Section 3 we present our algorithm. In Section 4 we present results of the applications of our method. In Section 5 we analyse the results obtained and compare them with other methods. Finally in Section 6 we expose the main conclusions along with future research lines.

2. Related work

In this paper we present an object detection method that detects automatically where objects appear in the image. This work can be enclosed into the category of regions of interest detectors, which are generally capable of reproducing similar performance that the human would provide in locating elemental features in images.

Feature detection is often a low-level image processing operation, usually performed as the first operation on an image, and examines every pixel to see if there is a feature present at that pixel. There are several feature detection methods that can be divided into four main groups: edge detectors (with methods such as Canny or Sobel [4]), corner detectors (such as Harris [5] or SUSAN [13]), blob detectors (such as Laplacian of

Gaussians [10], or MSER [9]) and region of interest detectors (SURF [2] or Intensity Extrema-based Region Detector [10]). In our case we take advantage of the presence of valleys in certain types of images, in order to develop our object detection method.

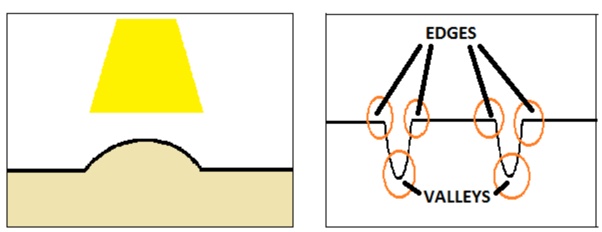
As it was mentioned in Section 1, object detection is often the first step in an object recognition method so, once the regions of interest are located in the image, we need to describe them in a way such the posterior object classification can give good results. There are many types of descriptors, usually divided according to the type of feature we are interested, such as shape, color or texture.

After the description of the detected objects is done, object classification is performed. It usually involves a learning procedure to find out which combination of features describe better an object so, when a new object arrives into the processing pipeline, it can be classified accurately. As it can be thought, having a three parts schema implies that the output of an stage (in this case the object detection stage) affects the results of the following stage hence the important of a good object detection, which is the objective of the work presented in this paper.

3. Methodology

3.1. Depth of valleys

As mentioned in Section 1, objects tend to appear as enclosed by intensity valleys when the source of light and the camera share the same or opposite direction (see Figure 2).



(a) (b)

Figure 2. Simulation of (a) an illuminated prominent surface (b) grey scale profile [3].

As it is explained in [8], ridges and valleys in bi-dimensional images are commonly identified as loci of minimum gradient magnitude along the relief's level curves. In 2D, ridges/valleys can be also identified as positive maxima/ negative minima of the curvature of the relief's level curves. Maxima are connected from one level to the next therefore constituting a subset of the vertex curves. As it is shown in Figure 3 the output of a ridges and valleys detector applied in one of our target images informs us about where valleys are in the image but not about their intensity. The results of these detectors are also affected by the presence of reflections or artifacts in the image.

In order to obtain information about the intensity of the valleys morphological gradient can be used. It is defined as the difference between the dilation and the erosion of

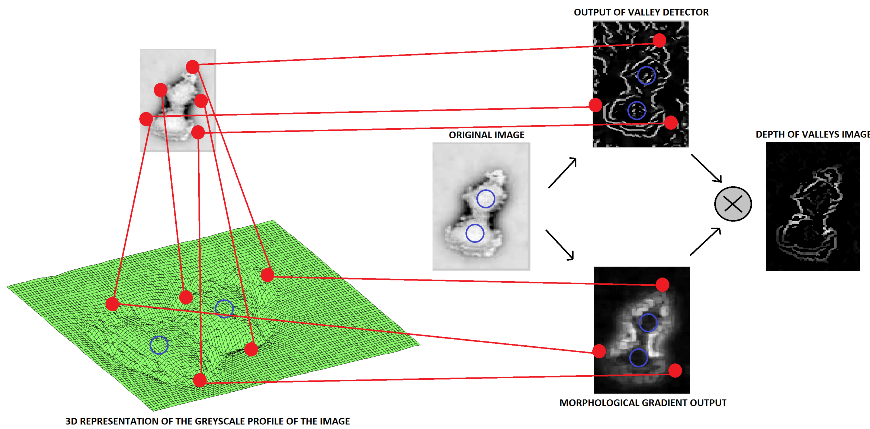


Figure 3. Example of the calculation of a depth of valleys image.

a given image [12] and gives as output an image where each pixel value indicates the contrast in intensity in the close neighborhood of that pixel.

As it can be seen in Figure 3, the combination of the information obtained from the ridges and valleys detector (*V*) with the one that the morphological gradient (*MG*) provides (see equation 1) gives as a result what we could call as the *depth of valleys image* (*DV*). In the points where we have a valley and the morphological gradient value is high (red points in Figure 3), we will have a high value of the deep of valleys image but in the points where we have a valley but the morphological gradient is slow, the depth of valley will be minima (blue circles in Figure 3).

$$\forall i, j \in \mathbf{I} \qquad DV(i, j) = V(i, j) \cdot MG(i, j) \tag{1}$$

We can see in Figure 3 how the depth of valleys has higher values in the points that constitute the 'real' valleys of the image and lower to points inside the valley.

It is important to mention that the ridges and valley extractor that we have used [8] needs of two parameter values (σ_d: differentiation scale and σ_i: integration scale) that must be put in correspondence with the size of the structural element (sd) that we use (in our case, a disk) to calculate the morphological gradient. More precisely, σ_i and sd should be equal in order to work in the same scale. If this does not happen, maxima points of the ridges and valleys extractor can be located in places where the morphological gradient is not maxima and therefore the desirable properties of the resulting depth of valleys image would be lost.

The next step consist of finding which parts of the image constitute the individual objects on the image, which will be separated by boundaries constituted by points with high values in the depth of valleys image.

3.2. Ellipse fitting depth of valleys accumulation algorithm

EF-DOVA algorithm consist of five different steps:

1. *Choice of the starting point:* The possibilities are either to use as starting points the minima of the depth of valleys image or the minima of the ridges and valleys image. Once

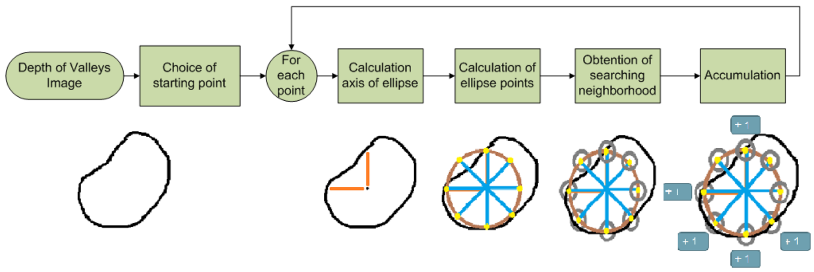


Figure 4. Graphical explanation of the depth of valleys accumulation algorithm.

the algorithm knows which are those points it starts the process. We have approximated the objects to be detected as having ideally an elliptical shape (although they do not appear as perfect ellipses in the majority of images) and our approach consists of trying to find the points in the image that fit better depth of valleys ellipses.

2. *Calculation of the axis of the ellipse:* For each starting point we find the maxima of depth of valleys image in both x and y direction. The positions of these maxima will constitute the axis of the ellipse that we want to approximate. Although the explanation is based in axis-centered ellipse, EF-DOVA algorithm also considers rotated ellipses.

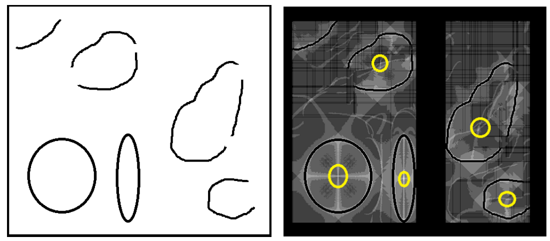
3. *Calculation of ellipse points:* In order to denote a point as center of the ellipse that has been approximated, it has to coincide with the depth of valleys image in a certain number of directions (that we can define as a threshold). In this case we consider 8 directions that go from $\Theta = 0$ to $\Theta = 360$ separated by 45 each point from another. In order to calculate the rest of the points we equal the equation of our approximated ellipse with the equation of the line with slope the tangent of each angle (eq. 2).

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1; \quad y = \tan\Theta \cdot x; x = \frac{a \cdot b}{\sqrt{a^2 \cdot m^2 + b^2}}; \ y = \tan\Theta \cdot \frac{a \cdot b}{\sqrt{a^2 \cdot m^2 + b^2}} \quad (2)$$

4. *Obtention of the searching neighborhood:* Once we have these 8 points we define a neighborhood around them where search for maxima in the depth of valleys image.

5. *Accumulation* To calculate the accumulation value for the point we count how many of the 8 ellipse points have maxima in their neighborhood.

The way the algorithm works can be better understood by observing in Figure 4 how it would work on an ideal binary depth of valleys image. Accumulation values go from 2 to 8. We will denote as high accumulation values those higher or equal than 5. In Figure 5 we can see how our algorithm works on a synthetic image that contains some of the shapes that we may find (with both closed and unclosed contours), some of them with even not a clear elliptical shape.



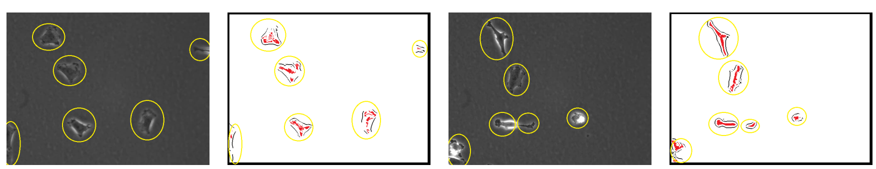
(a) (b)

Figure 5. (a) Original image (b) Accumulated values image.

4. Experimental results

4.1. Cell detection

With the increasing availability of live cell imaging technology, tracking cells and other moving objects in live cell videos has become a major challenge for bioinformatics [11]. One problem inherent for most tracking algorithms is over or under segmentation of cells, meaning that they tend to recognize one cell as several cells or vice-versa. Our approach is different than the one presented in [11] because they segment to identify cells (using flux tensors for detection of moving objects) while we do not need to do any segmentation to detect the different objects that appear in the image. As we show in Figure 6 we are able to detect which the cells are by only using our accumulation method, even when the intensity valley profile is far from closed. As it can be seen, we detect where cells are, even in cases where cells have just been divided (see Figure 6 (d).



(a) (b) (c) (d)

Figure 6. Example of cell identification (a,c) original images (b,d) binarized depth of valleys images (in black) with maxima of DOVA (in red).

We have measured, for the 361 frames of the video, the number of cells correctly detected and the number of false positives (objects detected that are not cells). In order to measure the accuracy of our test we have used the F_1 score, where $F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$. This measure considers both the precision and recall of the test to compute the score. The accuracy is better as the F_1 score is nearer to 1, meaning that, in the context of cell detection, we identify correctly a high number of regions that contain cells and in a very reduced number of cases we denote as a cell region one that does not contain a cell. In Figure 7 we show the results of F_1 score for our method and how they are affected by the scale parameter (value of σ_i and sd). As it can be seen, results are very good when we take values around 13 for the scale parameter.

Scale parameter	4	7	9	11	13	15
F1 score	0.839	0.878	0.88	0.893	0.898	0.8965

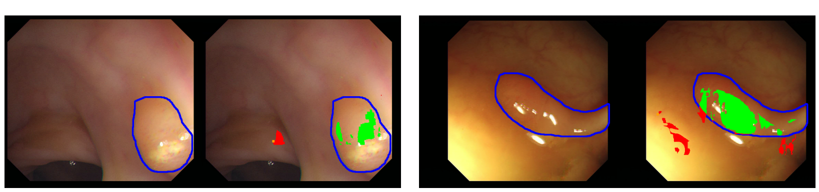
Figure 7. Results of F_1 score with respect to the scale parameter.

4.2. Polyp detection in colonoscopy videos

We have also tested our method in polyp detection in colonoscopy videos. Colonoscopy [6] is an screening technique used to detect colon cancer, which survival rate decreases as later it is detected hence the importance of an early detection. Many approaches have been carried in order to detect polyps in colonoscopy videos [1]. One of the main problem associated with this type of images is the great variability in polyp appearance (going from flat to peduncular shapes, or the view, from lateral to overhead) which makes unfeasible common template matching algorithms, among others.

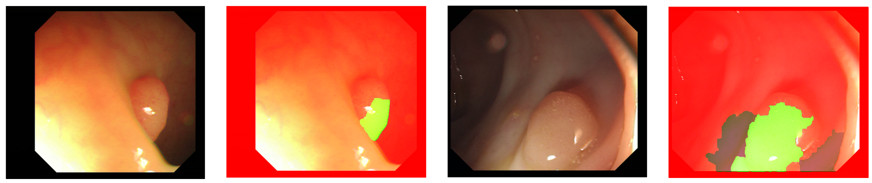
We show here how our method can help on reducing the part of the image where search for polyps. As it can be seen in Figure 8, our method locates maxima of accumulated value inside the polyp region (which are marked in green), no matter the type of polyp or the point of view of the image. It is true that it also places maxima in parts where is no polyp (marked in red), but our method offer an interesting result: if there is no maxima of accumulated value in some part of the image there will be no polyp in there, constituting a necessary but not sufficient condition of polyp presence.

In order to see how our method could help in polyp detection, we have colored areas in image according to the rate of pixels with high accumulated value rate in the area out of the total of points with high accumulated value, painting in reddish colors regions with low rate and in greenish colors those which high rate. The results (which can be seen in Figure 9 show that, although the type of polyp appearance in the images is different, our method colors in strong green parts of the image where the polyp really is. These results could be improved by using a more specific segmentation scheme.



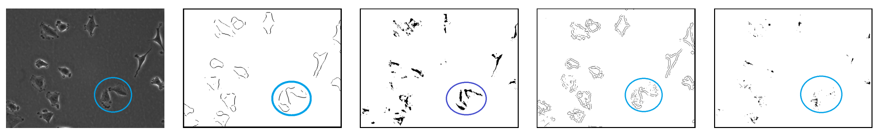
(a) (b)

Figure 8. Example of the apparition of maxima of accumulated value inside polyp regions.



(a) (b) (c) (d)

Figure 9. DOVA applied on polyp detection (a and c) original images (b and d) colored images.



(a) (b) (c) (d) (e)

Figure 10. Qualitative comparison between the results of DOVA and laplacian of gaussian (a) original image (b) binarized depth of valleys image (c) DOVA image (d) laplacian of gaussian image (e) accumulation of laplacian of gaussian.

5. Discussion

As shown in Section 4, our method offers good results in object region identification in several scenarios. The results could be improved if we adapt our method to specific problems (in the case of cell detection, we could study cues that indicate when cells are dividing) or we consider more kind of shapes to be approximated (although as of now, our method is able to cope with incomplete boundaries of any shape (see Figure 6).

We have compared the performance of our method with other approaches for the case of cell detection, where we have compared the results obtained with our complete cell detection algorithm with the results obtained by the combination of the Laplacian of Gaussians (LoG [10]) and our accumulation algorithm, to show the effect of using the DV image to obtain the boundaries of the objects (which can be seen as blobs as a whole) and to present how our accumulation algorithm can be used together with another feature detectors. In Figure 10 we show qualitative results of this comparison. For the sake of the comparison, we have compared the best result of our method with the best results achieved with LoG in terms of the objects' boundaries and applied the same parameters for the accumulation method (using LoG's output as the DV image). We can see that there are less maxima of accumulated value inside the cells for the case of LoG that, in some cases (see Figure 10 (d) may lead to lose the object.

In order to do a quantitative comparison we have measured, for frame of the video, the following: 1) percentage of the cells that have been correctly detected (cd; 2) number of false positives (parts of the image where we detects really non-existing objects); 3) percentage of pixels with high accumulated values near the center of the object; 4) the

	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5
DOVA	81.57%	12	26.55%	3.95	2.79
LoG + Accumulation	80.06%	3	22.84%	4.28	2.65

Table 1. Summary of the results. Experiment 1: percentage of correctly detected cells; Experiment 2: False positives; Experiment 3: Rate of high accumulation points near objects' center; Experiment 4: Mean accumulated value and Experiment 5: Ratio of high accumulation points versus low accumulation points per object.

mean accumulated value per object and 5) the ratio between pixels with high accumulated value vs pixels with zero as accumulated value per object. Results are shown in Table 1.

We can see the effect that has the use of DV image over LoG in object detection. In terms of object detection by applying our complete detection method, we detect more cells than when we apply the accumulation algorithm after LoG (81.57% against 80.06%, 30 cells) but, in the other hand, we present a higher total number of false positives (although the number is low). The reason why we do not have a perfect cell detection is that, in some cases of cell division, we detect one cell where there are really two. We can also see from Experiment 3 that by using EF-DOVA we obtain more pixels with high accumulated value near the real center of the object. In terms of mean accumulated value the results are very similar and we can see, from Experiment 5, that the ratio between the number of high accumulated values out of the total of pixels in the object is higher for DOVA. The main conclusion that can be extracted from this experiment is that the use of the DV image lead not only to a better cell detection but the number of pixels with high accumulated value near the center of the object is higher with our method, so we locate better the real center of the object.

The results of our method could be used, for example, to guide segmentation using high accumulation points as markers or, as a first step in an object recognition scheme where, once found which regions of the image contain objects, the next step often consists of describing them by using feature descriptors. The problem that we want to solve will tell us which type of feature descriptor we want to test (shape, texture or color).

The final objective in object recognition tasks is to classify the objects that appear in the image. To do so there are two main philosophies: generative (such as principal component analysis [14]) and discriminative methods (like support vector machines [7]). Generative approaches approximate original data to keep as much original information as possible whereas discriminative approaches are designed having classification tasks in mind. Classification goal is, given training data, to find the optimal decision criteria. In a generative model the likelihood of a sample is estimated and then assigned the most likely class while in discriminative models a label is assigned directly based on the decision criteria. As a classifier needs good inputs in order to perform well, we think the focus should be on object detection and description stages, in order to build accurate classification systems.

6. Conclusions and future work

In this paper a novel approach to object detection by object region identification has been presented, which takes advantage of the appearance of valleys surrounding objects. We have presented our two contributions: the definition of the depth of valleys image and the accumulation algorithm. Both are meant to work together although the accumulation algorithm could be applied, as it has been shown, to any kind of boundary image.

We have applied our method in several scenarios, showing promising results for different experiments (such as cell or polyp detection). In cell detection we show that our method is able to detect a great percentage of the cells that appear across all the images. In the case of polyp detection we show that by applying our method we offer a necessary but not sufficient condition of polyp presence.

Although our method could be improved if we adapt it to solve specific problems we have shown the strength and versatility of our method, even compared with other methods. Another future research line could be to consider other shapes in the approximation, such as squares or a more specific choice of the starting point in the accumulation. The results shown in this paper indicate that, considering all the possible modifications and improvements, our object detection method could be a powerful tool in object recognition (especially in images with the valley presence is clear).

ACKNOWLEDGEMENTS

This work was supported in part by a research grant from Universitat Autonòma de Barcelona 471-01-3/08, by the Spanish Government through the founded project ”COLON-QA” (TIN2009-10435) and by research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018).

References

[1] S. Ameling, S. Wirth, and D. Paulus. *Methods for Polyp Detection in Colonoscopy Videos: A Review*. Inst. fur Computervisualistik, 2009.
[2] H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. *Computer Vision–ECCV 2006*, pages 404–417, 2006.
[3] Bernal, J., Sánchez, J., and Vilariño, F. A Region Segmentation Method for Colonoscopy Images Using a Model of Polyp Appearance. In *Proceedings of the, IbPRIA 2011*, page (in press), Las Palmas de Gran Canaria, Spain, June 2011.
[4] R. Fisher, S. Perkins, and A. W. et al. Sobel operator.
[5] C. Harris and M. Stephens. A combined corner and edge detection. In *In Proceedings of The Fourth Alvey Vision Conference*, pages 147–151, 1988.
[6] Hassinger, J.P., Holubar, S.D. et al. Effectiveness of a Multimedia-Based Educational Intervention for Improving Colon Cancer Literacy in Screening Colonoscopy Patients. *Diseases of the Colon & Rectum*, 53(9):1301, 2010.
[7] M. Hearst, S. Dumais, and E. e. a. Osman. Support vector machines. *Intelligent Systems and their Applications, IEEE*, 13(4):18–28, 2002.
[8] López, A.M., Lumbreras, F. et al. Evaluation of methods for ridge and valley detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):327–335, 1999.
[9] J. Matas, O. Chum, M. Urbana, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.
[10] K. Mikolajczyk, T. Tuytelaars, and C. S. et al. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1):43–72, 2005.
[11] A. Mosig, S. Jager, and C. e. a. Wang. Tracking cells in Life Cell Imaging videos using topological alignments. *Algorithms for Molecular Biology*, 4(1):10, 2009.
[12] Serra, J. *Image Analysis and Mathematical Morphology*. Academic Press, Inc. Orlando, FL, USA, 1983.
[13] S. M. Smith and J. M. Brady. Susan - a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78, 1997.
[14] F. Wood, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometr. Intel. Lab. Syst*, 2:37–52, 1987.