Pattern Recognition in High-Throughput Zebrafish Imaging

Nezhinsky, A.E.

6 Discussion

6.1 Pattern Recognition in Computer Vision

In Chapters 2 and 3 we have presented three different approaches to retrieve a predefined biological shape from an input image. All approaches were based on deformable templates. The approaches were designed in such a way so that they could be used with in the HTS pipeline of zebrafish embryos. The methods needed to be able to retrieve shapes that are created as a result from different experiments, were subject to different light conditions and microscope settings. We have tried different representations of the zebrafish template.

First we have tried splitting the template into n vertical (to the zebrafish median axis) sub-templates that represent the boundary of the object. Subsequently the sub-templates were found one by one. As preprocessing step we have used an edge based representation. The edges were matched to each sub template. A mayor drawback was a high computational complexity due to a high amount of pixel comparisons within the image matrix space. This complexity could be reduced by applying a pyramid approach (lower resolution images) and introducing constraints to the location and orientation of the embryos within the image. One of the constraints that reduces computation time would be positioning the individuals with their head located left and their tail in the right part of the image in relation to the starting point. Since the zebrafish embryos differed in length, the size constraints of sub templates would complicate the analysis of differently shaped individuals.

Accordingly, we have tried using a variable number of sub templates, yet still with a fixed thickness. The prototype template that we proposed was represented as a bitmap that defines size limits for each sub template. It can easily be adapted to the needs of the application while the same algorithm is used.

Instead of looking at the edge data we considered filled contours. The same drawback as with the previous representation was that the shape was represented as a fixed maximal and minimal thickness at different shape locations, which excludes shapes that are slightly bigger or smaller from being matched.

We also introduced a shape normalization step. This was done by a straightening of the retrieved deformed shape based on the prototype template shape. This normalization step enabled comparison of differently deformed zebrafish embryos. However, due to the fixed thickness of sub-templates spatial analysis within individuals that differed in length were still difficult to compare.

In the third approach we represented the individuals based on their characteristic regions, rather than dividing into sub templates. This allowed for a better recognition of the individuals. The regions were found based on anchor points and during the search a higher order representation was used instead of pixel data. In this manner the computation time was reduced significantly and the problem of connected shapes was solved. Additionally, the constraint of locating embryos from left to right was lifted. Since the regions are recognized it is possible to do a more in depth analysis on certain regions of interest and compare the different shapes regardless of their deformation or length difference.

In order to localize a deformed template within the image space we examined a Genetic

Algorithm and a Dynamic Programming approach. First choice was the Genetic Algorithm due to its possibility to search for an optimal solution in large search spaces. This approach was feasible, yet in some cases algorithm could be trapped in a local optimum and was fairly slow due to high number of variables that were adjusted. In the third approach straightforward depth first graph searching could be used. This was motivated by the fact that the search graphs were relatively simple and thus no optimization was required. In cases where the graph is more complex a Dynamic Programming or another similar approach can be applied; such could be the case when images contain a large amount of different shapes.

The final version of the algorithm we propose does not rely on initial localization of the shape and therefore does not require any manual intervention or analysis. The framework can be easily adapted to work with other shapes, in the life sciences or in other fields that require accurate and robust shape retrieval. Further analysis of the object and the straightening thereof is part of future work.

A comparison of the result for the same image using the three algorithms is given in Figure 6.1.

6.2 Software Development

We have developed *ZFA*, a software solution that includes the interface and a processing unit, based on the algorithm that is presented in Chapter 3 (Approach 3).

Our solution is successful at recognizing the zebrafish larva shape and analyzing the infection within the larvae and is robust as it can be adapted to recognizing other shapes. The measurements that are performed on the infection are: localization, infection cluster size, infection cluster amount and average infection cluster intensity.

ZFA has proven its value as a software solution for the analysis in a HT screening pipeline.

6.3 Statistical Analysis

We have used the framework that we have developed for automated granuloma cluster recognition in order to analyze the spatial distribution in zebrafish larvae. As a proof of concept we have analyzed the data for the zebrafish larva infected with the wild-type Mycobacterium marinum (MM) and some of its mutants.

As a first assessment we compared the behavior of granuloma of the wild-type bacteria to the behavior of the mutant 714 strain. Mutant 714 was chosen as it was one of the mutants that did not make the fish ill and this assessment was used as a proof of concept for further investigation of other mutants. A statistical analysis on the spread of bacteria was performed and information on the spread of granuloma was derived. Since the MM tends to make the fish more sick it makes sense that overall more granuloma clusters are found in the wild-type infected fish. However, if we look at the normalized spread of infection it behaves approximately the same; it either stays at the site of injection or it moves towards the head of the larva. For the Mycobacterium marinum it seems that

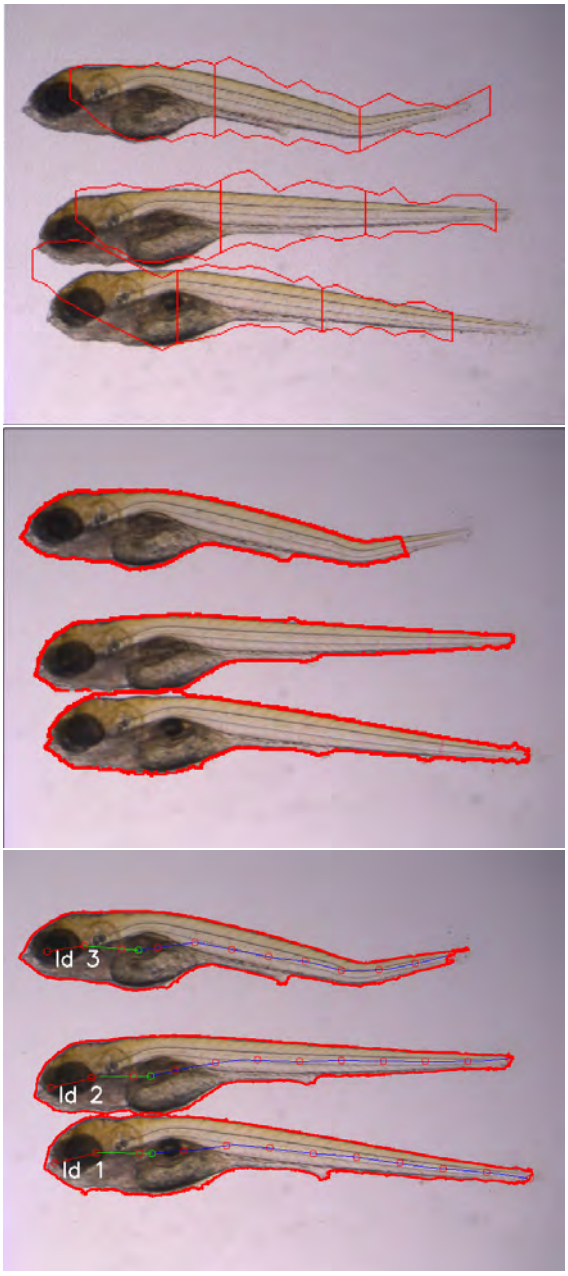

Figure 6.1: Shape retrieval results. The same input image is used for all three approaches (Approach 1 till 3 from top to bottom).

the infection is likely to migrate towards the head compared to the 714 mutant; in the 714 mutant it is established that the majority of the infection stays at the injection site.

In the second assessment we compared a large dataset of different mutant strains to the wild-type and to each other. Moreover, other measurement parameters could be considered in the analysis due to a more precise approach for retrieving the shape and its regions. Locations of interest such as the heart region and the injection point region could be retrieved. However, not all the mutants could be directly compared to the wild-type as they were not normally distributed. Those that were have been compared to each other and to the wild-type. Detailed results are provided in Section 5.4.

6.4 Conclusion

Here we present our final conclusions regarding our study of automation for High Throughput at the organismal level. We have applied our method to an important model organism, the zebrafish. With the development of our approach we made it possible to gain useful biological knowledge that used to be gathered manually in a High throughput fashion, which makes it a key activity within the field of bio-informatics.

We have shown and compared different deformable template based approaches that have been used for the pattern recognition step in the software. We have made it possible to automatically localize and annotate the shapes of the zebrafish larvae. The silhouettes that were identified serve as a mask to measure and analyze the fluorescent signal corresponding with the bacteria; if multiple fluorescent channels are used, other analysis can be accomplished with the same mask. In the case of zebrafish larvae, the current approach is sufficient for an analysis on spread through global regions in the zebrafish. However, in order to do an even more in depth analysis on infection spread, localization of certain organs within the fish or the determination of zebrafish age, we would require more characteristics than the silhouette. A silhouette representation does not suffice for that purpose since it is only describing the outside of the organism while inner information is left out. In this case an analysis of the brightfield signal within each mask can contribute with informative features. As an example we have experimented with a framework for the automated determination of the developmental stage of zebrafish embryos. A machine learning approach uses both silhouette and brightfield data from within the mask for feature extraction. In a preliminary setup we have set up a machine learning pipeline taking into account all signal in the brightfield image; the results are promising and indicate the feasibility of the approach (cf. Nezhinsky et al. [41], Figure 6.2).

We have discussed the construction and the evaluation of a workflow which we have embedded in a user interface to support the work with the algorithms described in this thesis in a high-throughput setting. So, during the development process prototypes of our software were already used for a real life problem. This evolutionary approach turned out to be successful. The reason for this successful design lies in the fact that evaluation and requirements were obtained from the intended user group which was involved in the development from the beginning and who were the intended end-users.

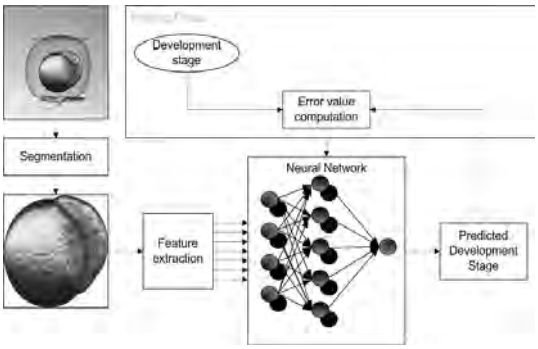

Figure 6.2: Overview of the proposed framework for automatic zebrafish embryo developmental stage determination.

The application of our software is described with a real life case study involving *Mycobacterium marinum* infection. The resulting data was analyzed for infection patterns. These patterns were related to the annotated areas within the zebrafish and therefore previously this kind of analysis was hampered due to the limitations given by manual analysis. With our algorithms the images could be processed in a High Throughput fashion and thus large amount of features could be generated from the input images. Using statistics on these features we were able to discover interesting differences in the organization and amounts of granuloma clusters for different bacterial mutants. By obtaining these results we have shown the need for such automation processes and smart software for these kind of screening applications.

Bibliography

[1] T. Bäck, Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms, Oxford Univ. press, 1996.

[2] B. A. Barut, L. I. Zon, Realizing the Potential of Zebrafish as a Model for Human Disease, Physiological Genomics 2 (2000) 49–51.

[3] J. S. Blackburn, S. Liu, A. R. Raimondi, M. S. Ignatius, C. D. Salthouse, D. M. Langenau, High-throughput imaging of adult fluorescent zebrafish with an LED fluorescence macroscope, Nature Protocols 6 (2011) 229–241.

[4] E. Borestein, E. Sharon, S. Ullman, Combining Top-down and Bottom-up Segmentation, Presceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop on Perceptual Organization in Computer Vision (2004) 46.

[5] G. Borgefors, Distance Transformations in Digital Images, Computer Vision, Graphics, and Image Processing 34/3 (1986) 344–371.

[6] G. Borgefors, R. Strand, An Approximation of the Maximal Inscribed Convex Set of a Digital Image, ICAIP (2005) 438–445.

[7] P. J. H. Bronkorsta, M. J. T. Reinders, E. A. Hendriks, J. Grimbergen, R. M. Heethaar, G. J. Brakenhoff, On-line Detection of Red Blood Cell Shape using Deformable Templates, Brakenhoff Pattern Recognition Letters 21(5) (2000) 413–424.

[8] R. Carvalho, J. de Sonneville, O. W. Stockhammer, N. D. L. Savage, W. J. Veneman, T. H. M. Ottenhoff, R. P. Dirks, A. H. Meijer, H. P. Spaink, A High-Throughput Screen for Tuberculosis Progression, PLoS One 6(2).

[9] T. F. Cootes, C. J. Taylor, D. H. Cooper, J. Graham, Active Shape Models - Their Training and Application, Comput. Vision and Image Understanding 61/1 (1995) 38–59.

[10] T. F. Cootes, C. J. Taylor, A. Lanitis, Active Shape Models: Evaluation of a Multi-Resolution Method for improving Image Search, Proceedings of the British Machine Vision Conference 1 (1994) 327–336.

[11] C. L. Cosma, K. Klein, R. Kim, D. Beery, L. Ramakrishnan, Mycobacterium Marinum Erp is a Virulence Determinant required for Cell Wall Integrity and Intracellular Survival, Infection and immunity 74(6) (2006) 3125–3133.

[12] J. Cousty, G. Bertrand, L. Najman, M. Couprie, Watershed Cuts: Minimum Spanning Forests and the Drop of Water Principle, Transactions on Pattern Analysis and Machine Intelligence. 31 (2009) 1362–1374.

[13] O. Cuisenaire, B. Macq, Fast and Exact Signed Euclidean Distance Transformation with Linear Complexity, IEEE International Conference on Acoustics, Speech and Signal Processing 6 (1999) 3293–3296.

[14] P. Danielsson, Euclidean distance mapping, Comput. Graph. Image Process. 14 (1980) 227–248.

[15] R. Fabbri, L. D. F. Costa, J. C. Torelli, O. M. Bruno, 2D Euclidean Distance Transform Algorithms: A Comparative Survey, ACM Computing Surveys 40 (2008) 1–44.

[16] P. F. Felzenszwalb, Representation and Detection of Shapes in Images, Ph.D. thesis, MIT, 2003.

[17] N. Funobiki, M. Isogai, An Eye-Contour Extraction Algorithm from Face Image using Deformable Template Matching, Mem. of the Fac. of Eng. 40 (2006) 78–87.

[18] A. Garrido, N. P. de la Blanca, Applying Deformable Templates for Cell Image Segmentation., Pattern Recognition 33/5 (2000) 821–832.

[19] S. George, T. Xia, R. Rallo, Y. Zhao, Z. Ji, S. Lin, X. Wang, H. Zhang, B. France, D. Schoenfeld, R. Damoiseaux, R. Liu, S. Lin, K. A. Bradley, Y. Cohen, A. E. Nel, Use of a High-Throughput Screening Approach Coupled with In Vivo Zebrafish Embryo Screening To Develop Hazard Ranking for Engineered Nanomaterials, ACS Nano 5/3 (2011) 1805–1817.

[20] R. Gerlai, High-Throughput Behavioral Screens: the First Step towards Finding Genes involved in Vertebrate Brain Function using Zebrafish, Molecules 15.4 (2010) 2609–2622.

[21] R. Gonzales, R. Woods, Digital Image Processing, Addison-Wesley, London, 2nd edn., 2001.

[22] X. He, R. S. Zemel, D. Ray, Learning and Incorporating Top-Down Cues in Image Segmentation, Lecture Notes in Computer Science 3951 (2006) 338–351.

[23] A. K. Jain, Y. Zhong, M. P. Dubuisson-Jolly, Deformable Template Models: A Review, Signal Processing 71/2 (1998) 109–129.

[24] A. K. Jain, Y. Zhong, S. Lakshmanan, Object Matching Using Deformable Templates, IEEE Transactions on Pattern Analysis and Machine Intelligence 18/3 (1996) 267–278.

[25] C. A. Karlovich, R. M. John, L. Ramirez, D. Y. Stainier, R. M. Myers, Characterization of the Huntington's Disease (HD) Gene Homologue in the Zebrafish Danio rerio, Gene 217(1-2) (1998) 117–25.

[26] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active Contour Models, International Journal of Computer Vision 1/4 (1987) 321–331.

[27] C. Kervrann, F. Heitz, A Hieraerchial Statical Framework for the Segmentation of Deformable Objects in Image Sequences, IEEE Computer Vision Pattern Recognition (1994) 724–728.

[28] H. Y. Kim, S. A. de Araujo, Grayscale Template-Matching Invariant to Rotation, Scale, Translation, Brightness and Contrast, IEEE Pacific-Rim Symposium on Image and Video Technology, Lecture Notes in Computer Science 4872 (2007) 100–113.

[29] F. Krolupper, J. Flusser, Polygonal Shape Detection for Recognition of Partially Occluded Objects, Pattern Recognition Letters 28 (2007) 1002–1011.

[30] B. Leroy, I. Herlin, L. Cohen, Multi-resolution Algorithms for Active Contour Models, ICAOS (1996) 58–65.

[31] A. Levin, Y. Weiss, Learning to Combine Bottom-Up and Top-Down Segmentation, Lecture Notes in Computer Science 3954 (2006) 581–594.

[32] J. P. Lewis, Fast Template Matching, Vision Interface 95 (1995) 120–123.

[33] L. Liu, S. Sclaroff, Region Segmentation via Deformable Model-Guided Split and Merge, Boston University Computer Science Technical Report 241 (2000) 98–104.

[34] R. Liu, S. Lin, R. Rallo, Y. Zhao, R. Damoiseaux, T. Xia, S. Lin, A. Nel, Y. Cohen, Automated Phenotype Recognition for Zebrafish Embryo Based In Vivo High Throughput Toxicity Screening of Engineered Nano-Materials, PLoS ONE 7(4) (2012) 1.

[35] M. R. Loken, Immunofluorescence Techniques in Flow Cytometry and Sorting, Wiley, 1990.

[36] D. R. Love, F. B. Pichler, A. Dodd, B. R. Copp, D. R. Greenwood, Technology for High-throughput Screens: the Present and Future using Zebrafish, Current opinion in biotechnology 15/6 (2004) 564–571.

[37] D. G. Lowe, Distinctive Image Features from Scale-invariant Keypoints, International Journal of Computer Vision 60(2) (2004) 91–110.

[38] L. Lucchesse, S. K. Mitra, Color Image Segmentation: A State-of-the-Art Survey, Proceedings of the Indian National Science Academy 67(2) (2001) 207–221.

[39] B. L. Miller, D. E. Goldberg, Genetic Algorithms, Tournament Selection, and the Effects of Noise, Complex Systems 9 (1995) 193–212.

[40] M. Monici, Cell and Tissue Autofluorescence Research and Diagnostic Applications, Biotechnol Annu. 11 (2005) 227–56.

[41] A. E. Nezhinsky, I. Martorelli, F. J. Verbeek., Detection of Developmental Stage in Zebrafish Embryos in a high throughput processing environment, Proceedings 20th Annual Belgian-Dutch Conference on Machine Learning (2011) 97/98.

[42] A. E. Nezhinsky, E. J. M. Stoop, A. A. Vasylevska, A. van der Sar, F. J. Verbeek, Spatial Analysis of Bacterial Infection Patterns in Zebrafish, Benelearn 21.

[43] A. E. Nezhinsky, F. J. Verbeek, Efficient and Robust Shape Retrieval from Deformable Templates, Leveraging Applications of Formal Methods, Verification and Validation. Applications and Case Studies, Lecture Notes in Computer Science 7610 (2012) 42–55.

[44] H. P. Ng, S. H. Ong, P. S. Goh, K. W. C. Foong, W. L. Nowinski., Template-based Automatic Segmentation of Masseter Using Prior Knowledge, IEEE Southwest Symposium on Image Analysis and Interpretation 2006 (2006) 208–212.

[45] R. Nohre, Deformed Template Matching by the Viterbi Algorithm, Pattern Recognition Letters 17/14 (1996) 1423–1428.

[46] N. Otsu, A Threshold Selection Method from Gray Level Histogram, IEEE Transactions on Systems Man Cybernet 9 (1975) 62–66.

[47] H. Peng, F. Long, X. Liu, S. K. Kim, E. W. Myers, Straightening Caenorhabditis elegans Images, Bioinformatics 24/2 (2008) 234–242.

[48] A. Ramirez, Circle Detection on Images using Genetic Algorithms, Pattern Recognition Letters 27/6 (2006) 652–657.

[49] C. B. Rohde, M. F. Yanik, Subcellular In vivo Time-lapse Imaging and Optical Manipulation of Caenorhabditis elegans in Standard Multiwell Plates, Nature Communications 2 (2011) 271.

[50] H. Samet, M. Tamminen, Efficient Component Labeling of Images of Arbitrary Dimension Represented by Linear Bintrees, IEEE Transactions on Pattern Analysis and Machine Intelligence 10/4 (1988) 579–586.

[51] J. Sethian, Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics Computer Vision and Material Sciences., Cambridge University Press, 1999.

[52] L. Shapiro, G. Stockman, Computer Vision, Prentice Hall, 2002.

[53] G. E. Sotak, K. L. Boyer, The Laplacian-of-Gaussian Kernel: a Formal Analysis and Design Procedure for Fast, Accurate Convolution and Full-frame Output, Computer Vision, Graphics, and Image Processing 48/2 (1989) 147–189.

[54] E. J. M. Stoop, T. Schipper, S. K. R. Huber, A. E. Nezhinsky, F. J. Verbeek, S. S. Gurcha, G. S. Besra, C. M. J. E. Vandenbroucke-Grauls, W. Bitter, A. M. van der Sar, Zebrafish Embryo Screen for Mycobacterial Genes Involved in Granuloma Formation reveals a Novel ESX-1 Component, Disease Model Mechanisms 4 (2011) 526–536.

[55] E. J. M. Stoop, Bacterial Factors involved in Tuberculosis Infection, Ph.D. thesis, VU, 2013.

[56] J. E. Sulston, E. Schierenberg, J. G. White, J. N. Thomson, The Embryonic Cell Lineage of the Nematode Caenorhabitis Elegans, Developmental Biology 100 (1983) 64–119.

[57] S. Svensson, G. Borgefors, I. Nystrom, On Reversible Skeletonization Using Anchor-Points from Distance Transform, Journal of Visual Communication and Image Representation 10 (1999) 379–397.

[58] H. D. Tagare, C. C. Jaffe, J. Duncan, Medical Image Databases a Content-based Retrieval Approach, Journal of the American Medical Informatics Association 4/3 (1997) 184–198.

[59] W. Tao, H. Jin, Y. Zhang, Color Image Segmentation based on Mean Shift and normalized Cuts, IEEE Transactions on Systems, Man, and Cybernetics 37/5 (2007) 1382–1389.

[60] E. Valveny, E. Marti, Deformable Template Matching within a Bayesian Framework for Handwritten Graphic Symbols Recognition, 3rd IAPR International Workshop on Graphics Recognition 1 (1999) 243–250.

[61] F. J. Verbeek, P. J. Boon, High-resolution 3D Reconstruction from Serial Sections: Microscope Instrumentation, Software Design, and its Implementations, International Symposium on Biomedical Optics. International Society for Optics and Photonics 4621 (2002) 65–76.

[62] A. Vogt, A. Cholewinski, X. Shen, S. G. Nelson, J. S. Lazo, M. Tsang, N. A. Hukriede, Automated Image-based Phenotypic Analysis in Zebrafish Embryos., Developmental dynamics 238/3 (2009) 656–663.

[63] H. E. Voikman, H. Clay, D. Beery, J. C. Chang, D. R. Sherman, L. Ramakrishnan, Tuberculous Granuloma Formation is enhanced by a Mycobacterium Virulence Determinant, PLoS biology 2/11 (2004) 1946.

[64] S. L. Walker, J. Ariga, J. R. Mathias, V. Coothankandaswamy, X. Xie, M. Distel, R. W. Koster, M. J. Parsons, K. N. Bhalla, M. T. Saxena, J. S. Mumm, Automated Reporter Quantification In Vivo: High-Throughput Screening Method for Reporter-Based Assays in Zebrafish, PLoS ONE 7(1) (2012) 1.

[65] Y. N. Wu, Z. Si, C. Fleming, S. C. Zhu, Deformable Template as Active Basis, International Conference on Computer Vision 11 (2007) 1–8.

[66] J. Yao, N. Kharma, P. Grogono, Fast Robust GA-Based Ellipse Detection, ICPR 2 (2004) 859–862.

[67] J. You, E. Pissaloux, W. P. Shu, H. A. Cohen, Efficient Image Matching: A Hierarchial Chamfer Matching Scheme via Distributed Systems, Real Time Imaging 1 (1995) 245–259.

[68] S. U. Yu, J. Shi, Object-Specific Figure-Ground Segregation, Computer Vision and Pattern Recognition (2003) 39–45.

[69] H. Zhao, J. Zhou, A. Robles-Kelly, J. Lu, J. Yang, Automatic Detection of Defective Zebrafish Embryos via Shape Analysis, Digital Image Computing: Techniques and Applications (2009) 431–438.

[70] Y. Zhong, A. K. Jain, Object Localization using Color, Texture and Shape, Pattern Recognition 33 (2000) 671–684.