FAST CONTENT-ADAPTIVE IMAGE SEGMENTATION

Theses of the PhD dissertation

Balázs Varga

Scientific advisor:
Kristóf Karacs PhD

Consultant:
Tamás Roska DSc
ordinary member of the
Hungarian Academy of Sciences

Pázmány Péter Catholic University
Faculty of Information Technology
Interdisciplinary Technical Sciences Doctoral School

Budapest, 2012

1. Introduction and Research Aims

The automation of tasks that can increase the quality, or extend the duration of human life has been under permanent and heavy research since the beginning of time. Such duties include, but are not restricted to jobs that are simply too monotonous (e.g. surveillance, or 24/7 quality assurance of mass produced items moving on a conveyor belt) or cannot be done by humans because of biological reasons (e.g. flight or automotive navigation tasks or medical imaging). In many of these problems we use visual data partially or exclusively, thus the accurate processing of visual information is inevitable for the refinement of the next generation of these machines. Of course, for the automation of complex human activities, machines and algorithms need to be equipped with a sensory-algorithmic arsenal that is somewhat similarly to the senses we humans possess. However, many of the more complex tasks that await to be mechanized are based not just on the processing, but on the ``understanding'' of visual information. The difficulty lies in the fact that besides the "sensory" input obtained by our eyes, the human brain uses complex cognitive information during the interpretation of the scene it sees. For example, we easily identify a tennis ball in case we have already seen one before. Looking at a bag of balls that have a similar color, we can grab one without any hesitation, even though the boundaries of neighboring balls may not be clearly visible due to poor lighting conditions. This is because we *know* something about the size and shape of such an object.

Interpreting this procedure in the language of image understanding, we can make two observations. The first is that high-level metadata (having *a priori* knowledge on the physical properties of a ball) can highly aid the accuracy of execution. On the other hand, a top-down approach can not succeed without using information provided by a complementary bottom-up processing (seeing the pile of balls, utilizing the retina channels to extract low-level information), because it is the fundamental basis which high-level information is applied upon. Being one of the most successful and straightforward sources to use, image understanding has always been inspired by the human visual system. As of today, many of the most successful algorithms in this field of computer vision utilize certain combinations of top-down and bottom-up approaches. Just as we know plenty about how the human visual system works and processes low-level visual information, we also know how to efficiently conduct certain image processing tasks from the pixel level, in a bottom-up manner. However, just like in the case of neurobiologists and psychologists who investigate the way how semantic information might be represented in the human brain, scientists in the field of computer vision have their own difficulties in finding an appropriate abstract representation and efficient application of top-down data.

My work aims at making a step towards efficient image understanding through the possibilities of fusing the top-down and the bottom-up approaches. In this dissertation the focus is put on the investigation of bottom-up image segmentation and the points where

injection of top-down knowledge into a data-driven segmenter is possible.

The first problem I have been working on was the fast segmentation of high resolution images. The motivation for using high resolution images is that this way we can obtain more information from the segmentation output, with objects or object details that otherwise could not be retrieved due to their small extent. Identification of a higher number of details can enhance the robustness of recognition and classification and it can also provide additional cues that top-down knowledge could be applied to. The downside of increasing the physical resolution, i.e. the number of pixels is that the amount of data to be processed grows, which has a negative impact on the running time. To overcome this problem, I constructed and implemented a framework that works in a data-parallel way, and therefore it can efficiently utilize the powerful computational capabilities of many-core environments. The principal idea of the system was inspired by the mean shift algorithm [6], that was extended with a quasi-random sampling scheme for further acceleration, and a cluster merging procedure to reduce over-segmentation. In addition, I proposed a method to reduce the overhead caused by parallelization.

The second problem I addressed was making an image sampling scheme adaptive, in order to take the local content of the image to be

segmented into account. (Observe Figure 1 as an illustration for the intuitive justification of the content-adaptive concept.)
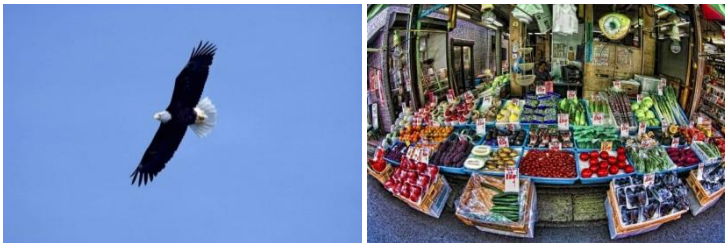
Figure 1. An intuitive example for the differences in content amount. Both images have a resolution of 14.5 megapixels, but while the one on the left contains a single object in front of a homogeneous background, the image on the right has far more details.

My goal was to keep the amount of time required for the segmentation to a content-dependent minimum, and at the same time to preserve an output quality similar to that of the naïve version, which is applied to every single input pixel without using any kind of sampling. To achieve this, I need not only to determine the number of samples required to maintain a certain segmentation quality, I also have to face the difficulty of finding a good spatial distribution of the samples. I developed an automated mechanism that is capable of solving this task. Additionally, the scheme uses a single-parameter for both the selection of sample candidates and for the registration of the strength of the bond between the pixels and their clusters. This way, the representation of the system remains compact, which enables the segmentation of large images as well. Additionally, this bond confidence strategy enables each pixel to be associated with the

most similar class, with respect to its spatial position and color. I designed the adaptive sampling method to fit into the realized parallel framework.

2. Methods of Investigation

For the design of the algorithmic background, I relied on the literature available on kernel density estimation, sampling theory, Gaussian mixture modeling, color space theory, similarity metrics and parallel algorithmic design.

For the first batch of evaluations, I considered three major analytical aspects that are most frequently taken into account for an extensive assessment of a segmentation framework. These are the following: *running time demand* (the amount of time required to provide the clustered output from the input image); *output accuracy* (can be measured using several different metrics that compare the output of our system to a ground truth); and *physical resolution* (equivalent to the number of input image pixels).

As one of my primary aims was to provide results that are comparable to the ones published in the literature, I used publicly available, well-known datasets [12][13][14] for the analysis of output quality. These databases have the advantage of providing a huge variety of standardized metrics (including Segmentation Covering [15], Probabilistic Rand Index [16], Variation of Information [17], F-measure [18], Average Precision [19], and Fragmentation [14]) in a unified evaluation framework.

However, these benchmarks contain images of relatively low resolution, therefore their applicability to the other two mentioned aspects is limited. Since the results measured on the datasets referred to above can not be extended in a straightforward manner onto images of higher resolution, I compiled two additional high resolution image sets, both containing real-life images taken in natural environmental conditions and depicting objects of various scales. The first set consists of 15 high quality images in five different resolutions (Table 1).

Name of extended graphics array	Abbreviation	Resolution	Resolution in megapixels (MP)
Wide Quad	WQXGA	2560×1600	4.1 MP
Wide Quad Super	WQSXGA	3200×2048	6.6 MP
Wide Quad Ultra	WQUXGA	3840×2400	9.2 MP
Hexadecatuple	HXGA	4096×3072	12.6 MP
Wide Hexadecatuple	WHXGA	5120×3200	16.4 MP

Table 1. Naming convention and resolution data of the images used for the measurements made with the parallelized system.

I used this set and a variety of general-purpose computing on graphics processing units (GPGPU) (Table 2) to assess the running time and the algorithmic scaling of my framework.

Device name	No. of stream processors	Clock frequency	Device memory	Compute capability
8800GT	112	1500 MHz	1024 MB	1.1
GTX280	240	1296 MHz	1024 MB	1.3
S1070SG	240	1440 MHz	4096 MB	1.3
C2050	448	1500 MHz	3072 MB	2.0
GTX580	512	1544 MHz	1536 MB	2.0

Table 2. The main parameters of the utilized general-purpose computing on graphics processing devices.

The evaluation made on this dataset confirmed my hypothesis that in the case of lossy, sampling-based segmentation algorithms, for a more complete evaluation a fourth analytical aspect, namely the *image content*, should be taken into account. Consequently, I composed a second image set of 103 images with each having a resolution of 10 megapixels. In the case of the measurements made using this set, the main dimension of evaluation was not how the alternation of resolution influences the running time, but how the varying amount of content does.

My framework was implemented in MATLAB [21] and I used the Jacket toolbox [22] developed by AccelerEyes. The statistical analysis was done using MATLAB and Microsoft Excel [23].

3. New Scientific Results

Thesis I. Parallelization and implementation of a bottom-up image segmentation algorithm on a many-core architecture.
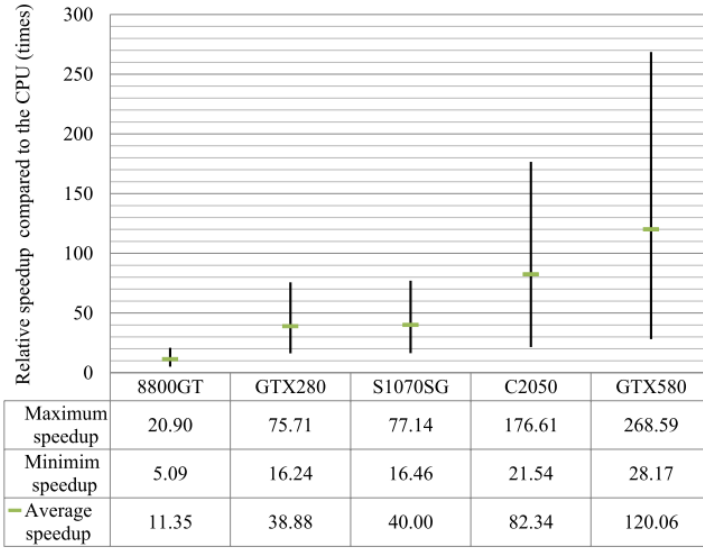
Most present-day photo sensors built into mainstream consumer cameras or even smartphones are capable of recording images of up to a dozen megapixels or more. In terms of computer vision tasks such as segmentation, image size is in most cases highly related to the running time of the algorithm. To maintain the same speed on increasingly large images, image processing algorithms have to run on increasingly powerful processing units. However, the traditional method of raising the core frequency to gain more speed – and computational throughput – has recently become limited due to the effect of high thermal dissipation, and the fact that semiconductor manufacturers are attacking atomic barriers in transistor design. For this reason, future development trends of different types of processing elements – such as digital signal processors, field programmable gate arrays or GPGPUs – point towards the development of multi-core and many-core processors that can face the challenge of computational hunger by utilizing multiple processing units simultaneously. However, these architectures require new algorithms that are not trivial to bring to effect.

Related publications of the Author: [2], [3], [4].

I.1. I parallelized the mean shift segmentation algorithm, which this way became capable of exploiting the extra computational power offered by many-core platforms. I applied the method to several different general-purpose computing on graphics processing devices and showed that the acceleration resulting from the parallelized structure is proportional to the number of stream processors.

I designed an image segmentation framework that performs mean shift iterations on multiple kernels simultaneously. By implementing the system on a many-core architecture and assessing it on multiple devices having various number of stream processors I have experimentally proven that the parallel algorithm works significantly faster than its sequential version, furthermore, raising the number of processing units results in additional acceleration.

Figure 2 displays the speedup of the mean shift core of the system compared to a CPU (Intel Core i7-920 processor clocked at 2.66GHz).

	8800GT	GTX280	S1070SG	C2050	GTX580
Maximum speedup	20.90	75.71	77.14	176.61	268.59
Minimim speedup	5.09	16.24	16.46	21.54	28.17
— Average speedup	11.35	38.88	40.00	82.34	120.06

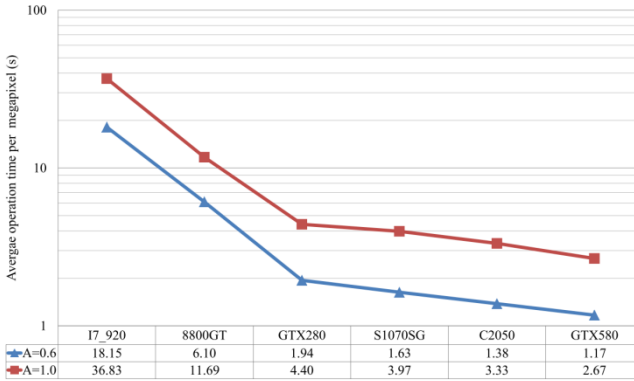**Figure 2.** Speedup results of the mean shift core obtained for different devices by pairwise comparison to the CPU.

I.2. **Through the analysis of the overhead caused by the parallel scheme I showed that by the early termination of kernels requiring remarkably more computations than the average, one can gain significant acceleration, while at the same time, segmentation accuracy hardly drops according to the metrics generally used in the literature.**

I found that it is not feasible to isolate saturated modes and replace them with new kernels in a "hot swap" way, due to the characteristics of block processing. I proposed a method (named abridging) to reduce the overhead caused by parallelization. I validated the relevance of the scheme through

quality (Figure 3) and running time evaluations made on the Berkeley Segmentation Dataset and Benchmark [12] and on a set of high resolution images (Figure 4).
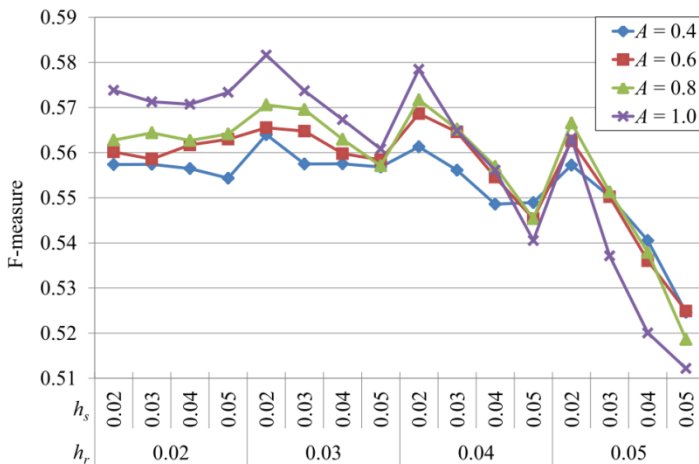
Figure 3. F-measure values obtained for the different parametrizations of the segmentation framework. h_s and h_r denote the spatial and range kernel bandwidths respectively, A is the abridging constant.

	I7_920	8800GT	GTX280	S1070SG	C2050	GTX580
A=0.6	18.15	6.10	1.94	1.63	1.38	1.17
A=1.0	36.83	11.69	4.40	3.97	3.33	2.67

Figure 4. Average running time of clustering one megapixel on different devices (and on CPU) as a function of the abridging parameter (*A*). The quality difference between the two settings is 3%.

I.3. I created an efficient, parallel cluster merging algorithm that can decrease the over-segmentation of segmented images by using color and topographic information.

The concept of over-segmentation is well-known [7][8] and widely used [9][10][11] in the image processing community. The main advantage of this scheme is that it makes the injection of both low and high-level information easy, thus the final cluster structure can be established using a set of rules that describe similarity with respect to the actual task. I have designed and implemented a parallel method for the computation of cluster neighborhood information and color similarity. Figure 5 gives an example of the results of the segmentation and merging procedures.

Figure 5. The input (left) is first segmented into clusters that share homogeneous characteristics (center), then clusters that are similar in terms of color and neighborhood consistency are merged (right).

Thesis II. Adaptive, image content-based sampling method for nonparametric image segmentation.

It is straightforward that the resolution of an image directly influences the running time and the output accuracy of a segmentation algorithm. But in case of lossy algorithms, the change in these two characteristics is not totally explained by the resolution because the distribution and amount of information in real-life images is very heterogeneous (see Figure 1), thus the results may depend on the characteristics of the input rather than the generic capabilities of the algorithm.

From the aspect of computational complexity, the obvious priority here is to minimize the number of samples, but simultaneously we have to keep in mind that undersampling introduces loss in image detail, whereas unnecessary over-sampling leads to computational overhead. To overcome these problems, I present the following contributions.

Related publication of the Author: [1].

II.1. I defined an implicitly calculated confidence value that is used as a heuristic for the adaptive sampling and at the same time is a sufficient guideline for the classification of image pixels.

I gave a single-parameter system that registers the strength of the bond between a pixel and the mode of a cluster, based on their spatial distance and color similarity. This way each picture element is associated with the class having the most similar characteristics. The key element for both the sampling procedure and the voting algorithm is the *bond confidence value* that is calculated implicitly during the segmentation without introducing any overhead.

II.2. I developed a sampling scheme guided by the content of the image that adaptively chooses new samples at the appropriate location in the course of the segmentation. By evaluating my framework on both my high resolution dataset and on publicly available segmentation databases, I verified numerically that the quality indicators of the adaptive procedure are almost identical to those of the naïve method (employed on all pixels) subject to all prevalent metrics, but at the same time the computational demand is remarkably lower.

My segmentation algorithm utilizes adaptive sampling such that the sampling frequency is based on local properties of the image. Homogeneous image regions get clustered fast,

initializing only a few large kernels, while spatially non-uniform regions, carrying fine details are processed using a larger number of smaller kernels that provide detailed information on them. While preserving the content of the image, this intelligent scheme reduces both the computational requirement and the memory demand, enabling the segmentation of high resolution images as well.

I showed via extensive output quality evaluation involving various metrics [14][15][16][17][18][19] on multiple datasets [12][13][14] that despite the fact that my algorithm uses sampling, the segmentation quality it provides fits in well among the publicly available alternatives built on the mean shift segmentation procedure.

I performed running time measurements on a set of 103 high resolution images. To cope with the lack of ground truth required for quality assessment, human subjects were asked to select the parametrization with which the best output quality of the most popular, publicly available reference segmenter [20] was obtained, and the parametrization of my framework that results in the most similar output to that of the reference. Running time results measured on the dataset using these settings are shown in Table 3.

	Mean running time (s)			Average speedup
	My system	Reference (High setting)	Reference (Naïve setting)	
Mean	**18.01**	320.01	23366.06	**18.58**
Std. dev./ (Relative)	**10.73** (59.59%)	277.46 (86.71%)	7385.67	**14.26**

Table 3. Running time results on the 103-item set of 10 megapixel images. The average speedup compares the running times of my framework with the reference system using the acceleration techniques it provides (high setting).

II.3. I created a measure to characterize the complexity of the content of images, and through high resolution time assessment and correlation analysis carried out between the running times and the amount of content indicated by my metric, I have empirically verified the adaptive behavior of my method, namely that the segmentation of images having less content is faster.

I defined a subjective, perception-based degree named the *kappa-index*. For a given image, it is calculated as the mean of ratings provided by human subjects, who are asked to assess the amount of useful content on a scale from 1 to 5, where 1 means a *"sparse image that contains only a few objects and large, homogenous regions"*, and 5 refers to a *"packed image having many identifiable details and rich information content"*. For each image in the 103-element high resolution dataset, the average rating of 15 participants was calculated and three subsets were formed based on the kappa-indices that represent the average amount of information in the images. The running

time results measured on the subsets are shown in Table 4 with *Class A* containing images with the least amount of information and *Class C* containing images with the most amount of information.

	Mean running time (s)			Average speedup	Average no. of kernels	No. of images
	My system	Reference (High setting)	Reference (Naïve setting)			
Class A	**12.85**	204.22	23273.47	**15.54**	599.43	30
Class B	**14.84**	344.63	23371.93	**22.24**	674.38	45
Class C	**28.62**	404.49	23455.82	**15.96**	1101.18	28

Table 4. Statistical results on the three subsets containing 10 megapixel images. Subsets are based on human ratings of the complexity of image content. The average speedup compares the running times of my framework with the reference system using the acceleration techniques it provides (high setting).

Measured on the whole high resolution image set, the correlation between the kappa-index and the number of kernels utilized per image by my algorithm is 0.694, which indicates that there is a strong connection between what human image annotators pointed out, and what my framework indicated as image content.

4. Application of the Results

The methods presented in my dissertation can be applied as an intermediate step in several different assignments of pattern recognition, object detection, and high-level image understanding. The segmentation algorithm has already been applied successfully for two real-life scenarios.

The first is the detection of crosswalks [4] that is a key task within the Bionic Eyeglass Project [24], an initiative that aims at giving personal assistance to the visually impaired in everyday tasks. The heart of this system is a portable device that, by analysing multimodal information (such as visual data, GPS coordinates and environmental noises), is able to identify and recognize several interesting object patterns (e.g. clothes and their colors, pictograms, banknotes or bus numbers) and situations (e.g. incoming bus at the bus stop, or environments such as home/street/office). The information provided by this sensorial input can be injected into my algorithm in the form of merging rules, among others. For example, if the system can localize the position of the user via GPS coordinates, the rule set of the segmenter can be extended dynamically to compose objects (such as lamp posts or signs) with particular (color or shape) properties. Detection robustness can be enhanced this way, as many false positives are filtered out.

As of now, smartphones offer not only a remarkable arsenal of sensors, but state of the art devices are also equipped with mobile GPUs that can be utilized by my parallel framework.

In the second case, my system was used in an early prototype of the Digital Holographic Microscope Project [25] that aims at developing an environment (hardware-software system) that is capable of autonomous water quality surveillance. To ensure robust detection, segmentation, and classification of different foreground objects (such as algae, pollens, or dust) from the background, the final version of the environment will use several different data sources including volumetric data (obtained via digital color holography) and material obtained with color and fluorescent microscopy. My framework was successfully used to segment the input provided in the form of a video frame sequence obtained using color light microscopy and fluorescent microscopy [26]. To ensure fast computation, the planned back-end system will be powered by specially designed many-core processors that again could be employed by my algorithm. Since the feature space my system works on is also a matter of selection, the use of additional channels gained by various sensors is a possible way to further improve in segmentation accuracy.

Acknowledgements

List of Publications

The Author's Journal Papers

[1] **B. Varga**, K. Karacs, "Towards a Balanced Trade-off Between Speed and Accuracy in Unsupervised Data-Driven Image Segmentation," sent to: *Machine Vision and Applications,* sent in May, 2012, under review

[2] **B. Varga**, K. Karacs, "High-resolution Image Segmentation using Fully Parallel Mean Shift," *EURASIP Journal on Advances in Signal Processing*, 2011:111 Springer, 2011

The Author's Conference Publications

[3] **B. Varga**, K. Karacs, "GPGPU Accelerated Scene Segmentation Using Nonparametric Clustering," *Proceedings of the International Symposium on Nonlinear Theory and its Applications (NOLTA)*, November 2010, Krakow, Poland, pp. 149–152

[4] M. Radványi, **B. Varga**, K. Karacs, "Advanced Crosswalk Detection for the Bionic Eyeglass", *Proceedings of the 12th International Workshop on Cellular Nanoscale Networks and Their Applications (CNNA)*, January 2010, Berkeley, CA, pp. 1–5

The Author's Other Publications

[5] H. van Welbergen, Z. Ruttkay, **B. Varga**, "Informed Use of Motion Synthesis Methods", Motion in Games, Springer Verlag, Berlin, ISBN 978-3-540-89219-9, 2008, pp. 132–143

References

Publications Related to Segmentation

[6] D. Comaniciu, P. Meer, "Mean Shift: a Robust Approach Toward Feature Space Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24 no. 5, 2002, pp. 603–619

[7] A. Duarte, Á. Sánchez, F. Fernández, A. Montemayor, "Improving Image Segmentation Quality Through Effective Region Merging using a Hierarchical Social Metaheuristic," *Pattern Recognition Letters* vol. 27 no. 11, 2006, pp. 1239–1251.

[8]   A. Hoover, G. Jean-Baptiste, X. Jiang, P.J. Flynn, H. Bunke, D.B. Goldgof, K. Bowyer, D.W. Eggert, A. Fitzgibbon, R.B. Fisher, "An Experimental Comparison of Range Image Segmentation Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18 no. 7, 1996, pp. 673–689

[9]   S. Beucher, "Watershed, Hierarchical Segmentation and Waterfall Algorithm," *Mathematical Morphology and its Applications to Image Processing*, 1994, pp. 69–76.

[10] L. Li, R. Socher, L. Fei-Fei, "Towards Total Scene Understanding: Classification, Annotation and Segmentation in an Automatic Framework," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2009, Miami, FL, pp. 2036–2043

[11] F. Shih, S. Cheng, "Automatic Seeded Region Growing for Color Image Segmentation," *Image and Vision Computing* vol. 23 no. 10, 2005, pp. 877–886.

Publications Related to Quality Evaluation

[12] D. R. Martin, C. C. Fowlkes, D. Tal, J. Malik, "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics," *Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV)*, vol. 2, 2001, pp. 416–423

[13] P. Arbeláez, M. Maire, C. C. Fowlkes, J. Malik, "Contour Detection and Hierarchical Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33 no. 5, 2011, pp. 898–916

[14] S. Alpert, M. Galun, A. Brandt, R. Basri, "Image Segmentation by Probabilistic Bottom-Up Aggregation and Cue Integration" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, February 2012, pp. 315–327

[15] M. Everingham, L. van Gool, C. Williams, J. Winn, A. Zisserman, "PASCAL 2008 Results," http://www.pascal-network.org/challenges /VOC/voc2008/workshop/index.html, 2008

[16] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward Objective Evaluation of Image Segmentation Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, 2007, pp. 929–944

[17] A. Y. Yang, J. Wright, Y. Ma, S. S. Sastry, "Unsupervised Segmentation of Natural Images via Lossy Data Compression,"

Computer Vision and Image Understanding, vol. 110, no. 2, 2008, pp. 212–225

[18] N. Chinchor, B. Sundheim, "MUC-5 Evaluation Metrics," *Proceedings of the 5th Conference on Message Understanding,* 1993, pp. 69–78

[19] B. Bartell, G. Cottrell, R. Belew, "Optimizing Parameters in a Ranked Retrieval System using Multi-query Relevance Feedback," *Proceedings of Symposium on Document Analysis and Information Retrieval (SDAIR)*, 1994

[20] C.M. Christoudias, B. Georgescu, P. Meer, "Synergism in Low Level Vision," *Proceedings of the 16th International IEEE Conference on Pattern Recognition*, vol. 4, August 2002, pp. 150–155

Publications Related to Algorithmic Development

[21] The MathWorks, Inc. MATLAB. *Addr.: The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA 01760-2098,* www.mathworks.com, accessed June, 2012

[22] AccelerEyes LLC, *Addr.: 800W Peachtree St NW, Atlanta, GA 30308*, www.accelereyes.com, accessed June, 2012

[23] Microsoft Corporation, *Addr.: 1 157th Avenue Northeast Redmond, WA 98052,* http://office.microsoft.com/en-us/excel/, accessed June, 2012

Publications Related to the Application of the Framework

[24] K. Karacs, A. Lázár, R. Wagner, D. Bálya, T. Roska, and M. Szuhaj, "Bionic Eyeglass: an Audio Guide for Visually Impaired," *Proc. of the First IEEE Biomedical Circuits and Systems Conference* (BIOCAS), London, UK, Dec. 2006, pp. 190–193.

[25] S. Tőkés, V. Szabó, L. Orzó, P. Divós, Z. Krivosija, "Digital Holographic Microscopy and CNN-based Image Processing for Biohazard Detection", *Proceedings of the 11th International Workshop on Nellular Neural Networks and Their Applications (CNNA),* July 2008, Santiago de Compostela, Spain, pp. 8

[26] B. Varga, "Color-based Object Segmentation for Water Quality Surveillance," *Thesis for Informatics Specialist in Bionic Computing (Post-graduate Specialist Training, Pázmány Péter Catholic University)*, 2011