LOCAL SEGMENTATION BY LARGE SCALE HYPOTHESIS TESTING

Segmentation as outlier detection

Sune Darkner, Anders B. Dahl, Rasmus Larsen
DTU Informatics, Technical University of Denmark, Richard Petersens plads, Kgs. Lyngby, Denmark
sda@imm.dtu.dk, abd@imm.dtu.dk,rl@imm.dtu.dk

Arnold Skimminge, Ellen Garde, Gunhild Waldemar
Danish Research Center for Magnetic Resonance, Copenhagen University Hospital Hvidovre, Kettegaard All 30 Hvidovre Denmark DK-2650
arnolds@drcmr.dk,elleng@drcmr.dk,Gunhild.Waldemar@rh.regionh.dk

Keywords: Segmentation, outlier detection, large scale hypothesis testing, locally adjusted threshold

Abstract: We propose a novel and efficient way of performing local image segmentation. For many applications a threshold of pixel intensities is sufficient but determine the appropriate threshold value can be difficult. In cases with large global intensity variation the threshold value has to be adapted locally. We propose a method based on large scale hypothesis testing with a consistent method for selecting an appropriate threshold for the given data. By estimating the background distribution we characterize the segment of interest as a set of outliers with a certain probability based on the estimated densities thus with what certainty the segmented object is not a part of the background. Because the method relies on local information it is very robust to changes in lighting conditions and shadowing effects. The method is applied to endoscopic images of small particles submerged in fluid captured through a microscope and we show how the method can handle transparent particles with significant glare point. The method generalizes to other problems. THis is illustrated by applying the method to camera calibration images and MRI of the midsagittal plane for gray and white matter separation and segmentation of the corpus callosum. Comparing the methods corpus callosum segmentation to manual segmentation an average dice score of 0.86 is obtained over 40 images.

1 INTRODUCTION

We present a novel way of performing binary segmentation of images with large global variations. In many segmentation problems a global threshold will not be appropriate, e.g. global changes in illumination, shadows, or background variations. These variations in pixel intensities can result in large segmentation errors if one threshold value is applied. As a consequence the threshold has to be locally adapted. Another problem is the dominating background intensities, which makes typical histogram based methods like histogram clustering (Otsu, 1975) inappropriate.

We propose a method based on the assumption that a local threshold exists, which will separate the segment of interest from the background. We present a well defined way of selecting the appropriate threshold value given the observations based on a large scale hypothesis test and experimentally show that this assumption is appropriate for many real segmentation problems.

2 PREVIOUS WORK

There exist a huge array of methods for classification and segmentation ranging from thresholds, classification, supervised like support vector machines (Boser et al., 1992), the Potts model (Potts, 1952), linear and quadratic discriminant analysis, regression and boosting and unsupervised methods such as mixture of gaussian. (Hastie et al., 2001) gives a good overview of these methods. More advanced methods such as level sets (Osher and Sethian, 1988), active shape and appearance models (Cootes and Taylor, 1994) and image registration (Modersitski, 2004). The method presented here is basically an image based threshold selection with the possibility to select a range rather than a just a threshold. An exhaustive review of threshold methods can be found in (Sezgin and Sankur, 2004).

Efron proposed in (Efron, 2004) to estimate the empirical null density function from the data and use this to identify cases of interest. The approach has been

used for genome responses in drug therapy and for identifying changes in shape (Darkner et al., 2007). the

3 LARGE SCALE HYPOTHESIS TESTING

The usual point of large-scale testing is to identify a small percentage of interesting cases that deserve further investigation using parametric modeling. The problem is that a part of the interesting cases may be extracted, but if more are wanted then also an unacceptably many false discoveries are identified (Efron, 2004). A major point of employing large-scale estimation methods is that they facilitate the estimation of the empirical null density rather than using the theoretical density. The empirical null may be considerably more dispersed than the usual theoretical null distribution. Besides from the selection of the non-null cases (the selection problem) large-scale testing also provides information of measuring the effectiveness of the test procedure (estimation problem). In this paper we employ both measures to separate the particles from the background transform calibration images into binary images, segment the corpus callosum and separate white and gray matter. Simultaneous hypothesis testing is founded on a set of N null hypotheses $\{H_i\}_{i=1}^N$ and test statistics which are possibly not independent. $\{Y_i\}_{i=1}^N$ and their associated p-values $\{P_i\}_{i=1}^N$ defining how strongly the observed value of Y_i contradicts H_i.

3.1 False discovery rate

In this paper we assume the N cases are divided into two classes, Null (background) and non-null (the particle) occurring with prior probabilities p_0 and $p_1 = 1 - p_0$. We denote the density of the test statistics given its class $f_0(z)$ and $f_1(z)$ (null or non-null respectively). False discovery rate methods are central to some large scale method and is employed here. It is typical to consider the actual distribution as a mixture of outcomes under the null and alternative hypotheses. Assumptions about the alternative hypothesis may be required. The sub-densities

$$f_0^+(z) = p_0 f_0(z) \quad , f_1^+(z) = p_1 f_1(z) \tag{1}$$

and mixture density

$$f(z) = f_0^+(z) + f_1^+(z) \tag{2}$$

leads directly to the local false discovery rate:

$$fdr(z) \equiv P(null|z_i = z)$$
$$= p_0 f_0(z)/f(z) = f_0^+(z)/f(z) \tag{3}$$

The false discovery rate describes the expected proportion of false positive results among all rejected null hypotheses and guarantees that the fraction of the number of false positives over the number of tests in which the null hypothesis was rejected (Efron, 2004). Figure 1 illustrates the fundamentals of the approach.

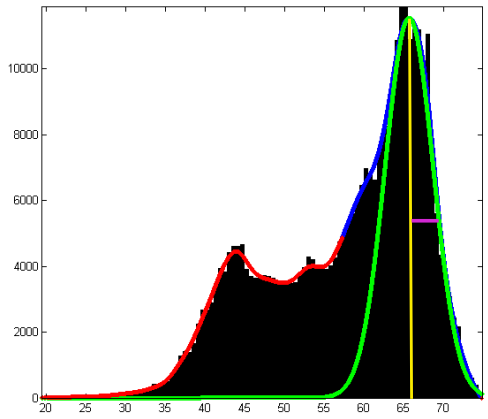


Figure 1: The graphical presentation of large scale hypothesis test. The red and blue curve is $f(z)$, the green is the pdf of the estimated null hypothesis $f_0^+(z)$ where the yello is the mean and the purple is the half width half maximum.

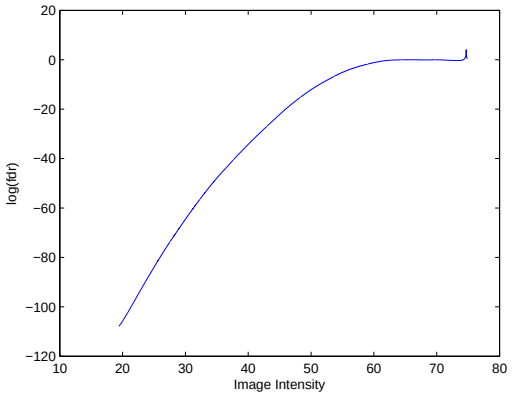


Figure 2: The corresponding logarithm of FDR for fig. 1.

For segmentation the hypothesis test is used to find pixels that are a part of a particle i.e. observations that deviates significantly from the average local background. We use large-scale testing to estimate the empirical null hypothesis for a given region assuming the pixel values follows some normal distribution.It is convenient to consider $z_i = \Phi^{-1}(P_i), \quad i = 1 \ldots N$ where Φ is the standard normal c.d.f and $z_i|h_i \sim N(0,1)$. Estimates of the pixel error and confidence bounds can be mapped back to the distribution of deformation vectors through Φ. Additional infor-

mation is that the background has higher pixel intensities than the particles. The background will therefor be the highest and largest distribution. This can be used to get a better empirical estimate of of the null hypothesis.

3.2 Estimation of H_0

Assuming that the intensities of the background follows a normal distribution and that it the background is brighter than the object of interest, due to the back lighting we are able to estimate H_0. Thus for the estimation we choose an appropriate resolution and map all value into the histogram. This forms our joint distribution $f(z)$ to which we then fit a spline with appropriate smoothness for an approximation of the joint distribution. We can then identify the first large peak as the mean value of $f_0^+(z)$ and use half width, half maximum to estimate the standard deviation. Now the obvious choice would be full width half maximum, however there is a blending of in the joint distribution of $f_0^+(z)$ and $f_1^+(z)$ which gives a thicker tail towards lower intensity values and a much more conservative estimate of the standard deviation. se figure 1.

3.3 Parameters and Their Interpretation

In practice there are several parameters to be selected. The first is the level at which we are willing to accept false positives which is the an expression of the certainty that a given observation is really different and not just a part of H_0. This does not in anyway tell us that the class is a certain kind of tissue or particle, just that this is with a certainty of say 99.9% different from the null distribution thus the observation is an outlier. The testing area has to be select. This criteria is mainly driven by the object in question and the background. Sufficient information about the distribution of the object and background must be present and on the other hand non-important features should be minimized. The spatial sampling density has to be selected as well. In all experiments in this paper we have up-sampled the image image by a factor of 10-100 which also yields an equal sub-pixel resolution of the method. Usually the test is based on $10^5 - 10^6$ samples and in a 100 bin histogram which ensures smooth nest of $f(z)$ and enough resolution for gray values. In practice $f(z)$ is approximated by a spline thus the smoothness has to be selected for good estimation of $f_0^+(z)$ and can compensate for low number of sample.

4 EXPERIMENTS

We have applied the method to 3 sets of data. Small particles obtained with high magnification, MRI of the brain for segmentation of the corpus callosa and a standard checker board for image calibration with highly varying intensity values. These 3 diverse applications show the versatility of this very simple but robust method. For all examples we have shown the sampling area in the sampling resolution for both segmentation and object.

4.1 Particles

For characterization of powders, droplets etc the size and shape of the objects are very important. In order to do a good classification the particles need to be found and properly segmented. The method is well suited for images where the light source changes in intensity and distribution locally and globally from frame to frame in a series of images, cases where robust estimate for background removal can be difficult to obtain. In addition some particles may partially shadow other particles thus making the global background removal incapable of segmenting the particle in question. The method has been tested on 3 types of particle images. A set of LED back lit particles suspended between 2 sheets of glass with varying distance to the focus plane . A tracking scenario with time series of calibrated particles suspended in water and some laser back lit non-uniform particles suspended in water.

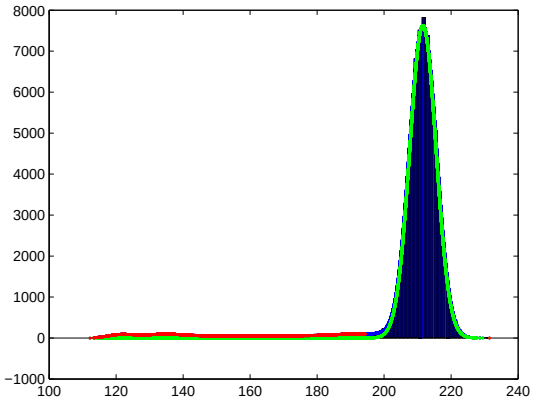
4.1.1 Fixed Particle

We have applied the method to the 25 μm particles suspended in water between two sheets of glass in with different distances to the focal plane. Figure 4 show a 25 μm particle at 6 distances to the focal plane. This experiment illustrates the sensitivity of the method, where even vaguely visible particles can be segmented without parameter tuning.

4.1.2 Shadowing effect

We have preprocessed all images with multi scale blob detection such that we have rough estimate of the size and location of the blobs. this is sufficient to perform the segmentation. The data consists of movie sequences obtained with 5 times magnification of semi transparent particles of 100 50 25 and 5 μm in a water solution and used for illustration of handling of shadowing effects without change to the parameters of the algorithm. Figure 5 show a segmentation performed over several frames where a larger particle passes in

(a) The particle (b) The segmentation



(c) The hypothesis test histogram

Figure 3: A typical segmentation and the corresponding histogram. The green line in the histogram is the estimated H_0 and the red line indicates outlier with the fdr=0.0001 The histogram clearly show how well our assumption of the background following a normal distribution holds. The outlier are the particle we segmenting.

the background creating a shadowing effect. The example illustrates how the methods can handle changes in the illumination without failure.

4.1.3 Real Particles

To show how the method handles real world data we have segmented some particles of some material samples from the industry. The chance of encountering nice uniform spherical particles in the real world is quite small except from droplets. Figure 6 and 7 show two small examples of such particles. The segmentation is very good due to the locally very uniform background, making the distribution of peaked and narrow i.e. a small standard deviation. Glare points are also handled very well but there is naturally enough a band around the glare point where the particle is misclassified. This is due to the fact that the method is just a simple threshold without any spatial prior. These gaps can be handled by applying morphological operations such as closing.

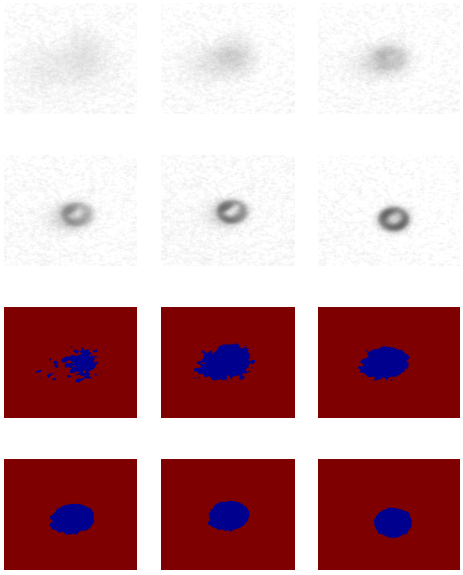


Figure 4: The images show the particle at 6 different distances to the focal plane, with the las image being of the particle in focus. Below are the segmentations of each od the images in the same order. The images show that the particle can be segmented even if the signal is very vague and the glare point is correctly classified as a part of the particle.

4.2 Calibration Images

We show some results here on calibration images. When using a well known structure as the checkerboard for calibration it is important to exactly locate the corner of the squares. This can be by morphological operations on a segmented image. The proposed method delivers very good segmentation that can easily be used to derive the exact sub-pixel position of the corners creating a robust foundation for image calibration. The figures 8 and 9 below show the results of the segmentation

4.3 MRI

To illustrate the method on another modality, we have applied the method to MRI scans of the human head the midsagittal plane that contain the subject corpus callosum (se figure 10). This data is a part of the LADIS study (Pantoni et al., 2005). The method have applied to white matter gray matter segmentation and segmentation of the corpus callosum. In the latter case we have a manually segmented truth, thus we can compute the segmentation error via the Dice coefficient (Sørensen,).

Across 40 subject with their Corpus Callosum segmented with the AAM (Ryberg et al., 2006) and manually corrected we found that local segmentation

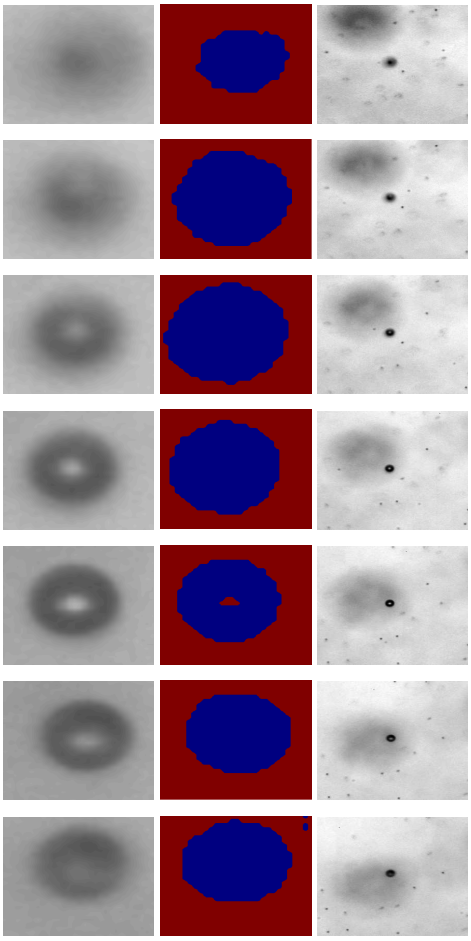
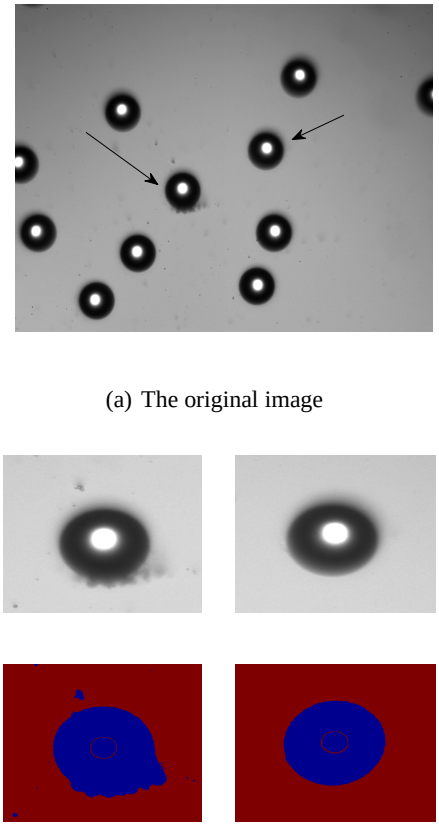


Figure 5: The figure show a segmentation performed with the same parameters on the same object subject to changing shadowing effect caused by a large particle passing in the background.

through large scale hypothesis testing gave an average Dice coefficient of 0.856 with *std* 0.034. The Corpus Callosum was extracted in the same hypothesis test, however if we use the local property and us a smaller window, outliers become more significant and we get better segmentation. By switching to a more local neighborhood we get a improvement of almost 3% to 0.88 and the difference is very significant $(p << 0.01)$

5 SUMMARY AND CONCLUSION

We have presented a local adaptive method for binary segmentation. The methods has successfully been tested on particle images for particle segmentation, calibration images for landmark extraction and and midsagittal slices of MRI for segmentation of corpus callosum and gray matter white matter segmentation. The method is very robust with respect to changes in intensity across the image and statistically characterizes the resulting segmentation. We have shown that compared to manual segmentation of the orpus callosum we can achieve a dice coefficient of 0.86 on using a mosaik of 5 patches. The method is directly extendable to 3D, other types of distribution as shown in (**?**). The hypothesis test and FDR should be extended to higher dimensions that the one dimensional case discussed here and tested on several other types of images. furthermore the algorithm should be implemented such that it can handle multiple classes and segment a whole image in one go.

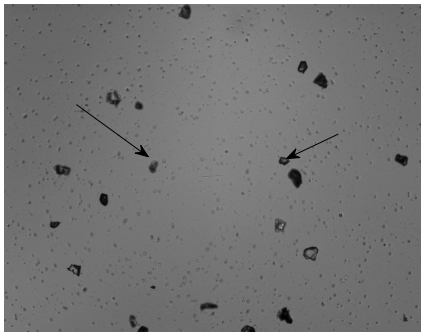


(a) The original image

Figure 6: The figure show some real world samples. The figure show that the segmentation the glare points is handled very well. The small 'gap' can be fixed by a simple morphological operation

REFERENCES

Boser, B., Guyon, I., and Vapnik, V. (1992). A training algorithm for optimal margin classifiers. *Fifth Annual Workshop on Computational Learning Theory*, pages 144–152.



(a) The original image

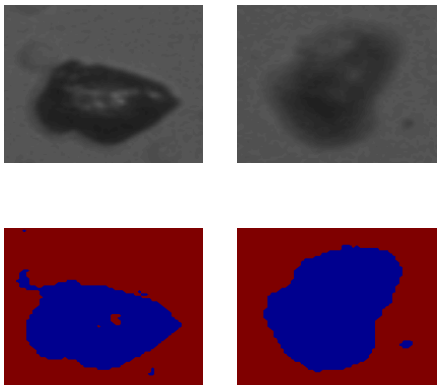


Figure 7: Some crystal like particles are shown in the figure. In spite of the relative low difference between background and object and the fact that the samples are semi transparant, the segmentation is good. Even small ones are handled together with the large ones.

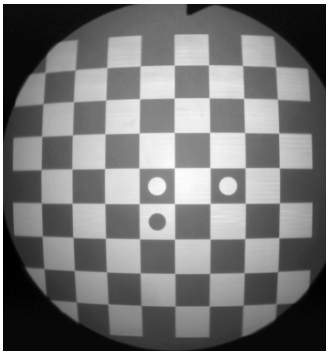


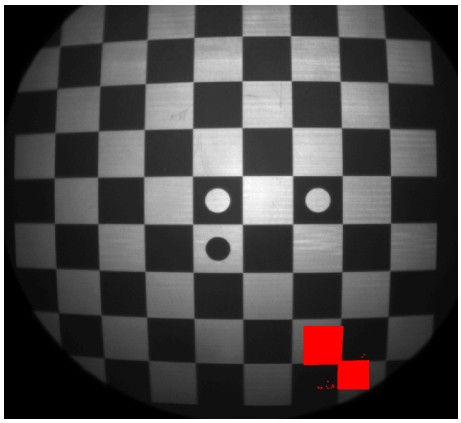
Figure 8: The original calibration image. As can be seen the intensities varies significantly with the highest values at the center and decaying radially.

Cootes, T. and Taylor, C. (1994). Using grey-level models to improve active shape model search. In *Proceedings of International Conference Pattern Recognition*, pages A:63–67.

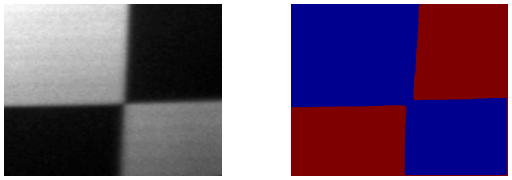
Darkner, S., Paulsen, R. R., and Larsen, R. (2007). Analysis of deformation of the human ear and canal caused by mandibular movement. In *Medical Image Computing and Computer Assisted Intervention MICCAI 2007*, pages 801–8, B. Brisbane, Australia, Springer Lecture Notes.

Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *Journal of the American Statistical Association*, 99(465):96–104.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag.



(a) The calibration image



(b) Image part segmented (c) Segmentation

Figure 9: The original calibration image. As can be seen the intensities varies significantly with the highest values at the center and decaying radially.

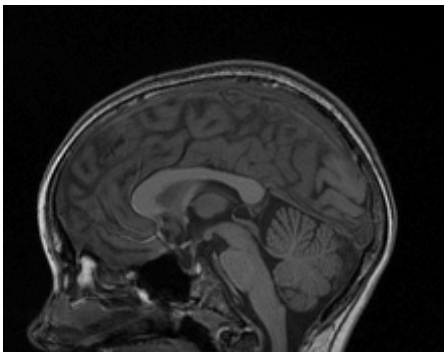


Figure 10: The midsagittal slice from an MRI of a head.

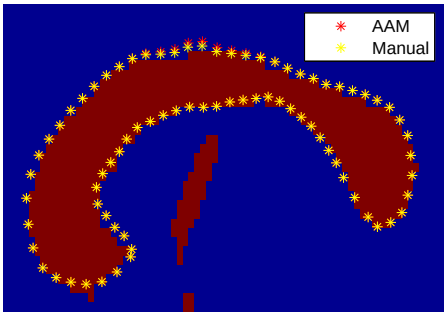
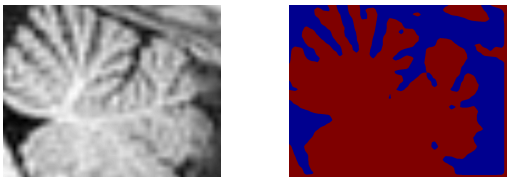
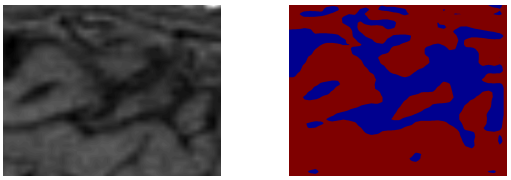


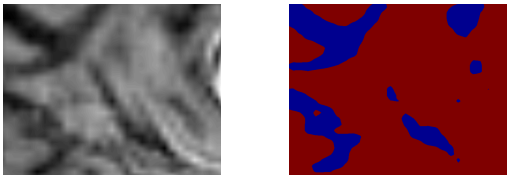
Figure 11: This figure show the result of the segmentation of fig 10 using the method proposed in this paper. The red dots are the segmentation achieved by the AAM and the yellow the manual segmentation. The read part of the image is the segmentation with the our method. This result show how efficient this algorithm is for local segmentation



(a) Image part segmented (b) segmentation



(c) Image part segmented (d) segmentation



(e) Image part segmented (f) segmentation

Figure 12: The original calibration image. As can be seen the intensities varies significantly with the highest values at the center and decaying radially.

Modersitski, J. (2004). *Numerical Methods for Image Registration*. Oxford University Press.

Osher, S. and Sethian, J. (1988). Fronts propagating with curvature dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79:12–49.

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, 11:285–296.

Pantoni, L., Basile, A., Pracucci, G., Asplund, K., Bogousslavsky, J., Chabriat, H., Erkinjuntti, T., Fazekas, F., Ferro, J., Hennerici, M., et al. (2005). Impact of age-related cerebral white matter changes on the transition to disability-the LADIS study: rationale, design and methodology. *Neuroepidemiology*, 24(1-2):51–62.

Potts, R. (1952). Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophical Society*, 48:106–109.

Ryberg, C., Stegmann, M. B., Sjöstrand, K., Rostrup, E., Barkhof, F., Fazekas, F., and Waldemar, G. (2006). Corpus callosum partitioning schemes and their effect on callosal morphometry.

Sezgin, M. and Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–168.

Sørensen, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biologiske Skrifter*.