Author's Accepted Manuscript

This is a post-peer-review, pre-copyedit version of an article published as:

Fuzzy Segmentation of the Left Ventricle in Cardiac MRI using Physiological Constraints

Tasos Papastylianou[1], Christopher Kelly[1], Benjamin Villard[1], Erica Dall' Armellina[2], and Vicente Grau[1]

[1] Institute of Biomedical Engineering, University of Oxford, Oxford, UK,
`tasos.papastylianou@kellogg.ox.ac.uk`,
WWW home page: `http://tpapastylianou.com`
[2] Acute Vascular Imaging Centre, John Radcliffe Hospital, Oxford, UK

Abstract. We describe a general framework for adapting existing segmentation algorithms, such that the need for optimisation of intrinsic, potentially unintuitive parameters is minimized, focusing instead on applying intuitive physiological constraints. This allows clinicians to easily influence existing tools of their choice towards outcomes with physiological properties that are more relevant to their particular clinical contexts, without having to deal with the optimisation specifics of a particular algorithm's intrinsic parameters. This is achieved by a structured exploration of the parameter space resulting in a subspace of relevant segmentations, and by subsequent fusion biased towards segmentations that best adhere to the imposed constraints. We demonstrate this technique on an algorithm used by a validated, and freely available cardiac segmentation suite (Segment – `http://segment.heiberg.se`).

Keywords: cineMRI,heart,probabilistic,segmentation

Abbreviations: *MRI* – Magnetic Resonance Imaging, *CT* – Computed Tomography, *SSFP* – Steady-State Free Precession, *LGE* – Late Gadolinium Enhancement, *EF* – Ejection Fraction, *SV* – Stroke Volume, *SA* – Short Axis, *LV* – Left Ventricle, *PCA* – Percutaneous Coronary Angioplasty, *MI* – Myocardial Infract

1 Introduction

Cardiac cineMRI is rapidly becoming one of the leading investigations in the assessment of cardiac disease, due to its ability to capture good quality images of the whole heart throughout the cardiac cycle, allowing for a more dynamic assessment of the heart. Any measurements relating to the ventricles that can be made from these images, such as stroke volume and ejection fraction, require segmenting the blood pool from the myocardium and background, for use in further calculations or processing. Segmentation of the ventricles by manual delineation from an expert clinician is prohibitively time-consuming, therefore much research has focused on automated techniques for segmentation.

A large number of approaches to cardiac segmentation from cardiac MRI (and other modalities) have been developed over the past few decades, such as approaches based on Deformable models or Shape models, Registration-based techniques, Image- / Voxel-based classification methods, or combinations of the above[1]. A common limitation of all these methods is that they tend to be reliant on a careful selection of parameters for optimal performance. However, such parameters tend to reflect more the algorithm's inner workings, rather than the clinical task at hand, and the choice of an optimal set of parameters can therefore be fairly unintuitive to the clinician.

Segmentation algorithms traditionally produce deterministic results (i.e. a clearly defined label per voxel); however, algorithms that produce fuzzy or probabilistic results are becoming increasingly common, as they have several advantages, such as being able to represent more complex situations relating to partial volume voxels, and therefore lead to more accurate estimation of clinical parameters. An example of this fuzzy approach is "ensemble" methods, such as Adaboost[2], where a stronger classifier is built from a weighted collection of weaker classifiers, or such as STAPLE[3] which attempts to produce a better result from an existing pool of segmentation results (where these are already considered to be of reasonable quality, i.e. derived from strong classifiers) via some form of weighted consensus.

1.1 Our approach

We propose a framework which aims to enhance existing segmentation algorithms, with the following goals in mind: 1) in the case of deterministic algorithms, propose a generalisable method to produce a probabilistic equivalent, by exploring the parameter space in a structured manner so as to produce a segmentation space, and then fusing the results appropriately using an ensemble approach; 2) reduce the need for predefined, optimal, problem-specific parameter sets, by weighing each segmentation in the segmentation space according to its compatibility with respect to intuitively defined physiological constraints, before fusing them together into a fuzzy segmentation result. 3) Use the above to improve the output and usability of *existing* algorithms from a *clinical* viewpoint, by allowing clinicians to guide segmentation algorithms towards results that are more relevant to their particular clinical context by simply defining such a context in intuitive clinical terms, rather than attempt algorithmic optimization via trial and error over an unintuitive set of parameters.

We demonstrate the framework on an existing segmentation algorithm proposed by Heiberg et al. [4]. This algorithm was chosen both because of its simplicity, making demonstration of the concept straightforward, and also because an implementation is made freely available online by the authors (Segment — http://segment.heiberg.se), which has been shown to be robust and produce accurate results (an extensive list of publications citing the algorithm is made available on the project website).

Paper organisation: The remainder of the paper is organised as follows: Section 2 provides a brief background for the Heiberg algorithm, which is used to demonstrate the proposed framework; Section 3 expands on the motivation behind our approach; Section 4 details the method used to obtain the fused segmentation; Section 5 presents the results with respect to clinician-provided manual segmentations used as a gold standard.

2   Background

The 2005 algorithm proposed by Heiberg et al. [4] (henceforth called the Heiberg algorithm) is essentially a *deformable model*-based segmentation approach. The model, consisting of a set of 2D active contours (one per slice), seeks to achieve an equilibrium between two competing sets of forces acting on its surface, while taking into account within-slice and temporal information; at each iteration, external forces guide the evolution of the model towards image-dependent features, whereas internal forces constrain the evolution, such that model smoothness and shape are relatively preserved.

The model has two external forces, an inflating *Balloon force*, and an *Edge force*. The Balloon force is dependent on local intensity, favouring expansion of the contour in areas closer to the estimated object's average intensity (as initialised by the user by selecting a single voxel lying within the left ventricle from the image). The Edge force is defined in terms of edge images derived from the image. Four edge images are produced, corresponding to estimating image edges in 4 different directions. At the point of calculation of the Edge force, the most appropriate edges to evolve towards are chosen given the direction of evolution of the model. Temporal information is introduced to the model by smoothing the edge force at each node-point of the model over several timeframes.

There are four internal forces with the purpose of ensuring spatial and temporal smoothness: a *Curvature force* which promotes smoothness in the overall contour shape, a *Damping force* and an *Acceleration force*, which ensure spatial continuity of the model's nodes within timeframes, and a *Slice force* which relatively discourages node movement between the slices (i.e. in the z-plane).

The above six forces are then combined in a "modality dependent" manner to control model evolution; here, "modality dependent" means choosing a set of modifiers for each force (i.e. the algorithm's parameters), which are most effective at leading the model towards a successful segmentation, given a particular investigation or image type.

Details of the mathematical implementation of these forces are beyond the scope of the present paper — particularly in the context of proposing a generalised framework aiming to minimize the role played by individual parameters, and by extension their particular role in the underlying mechanics of the algorithm in question; we refer the interested reader to the original paper for implementation details.

3   Motivation

As with most segmentation algorithms, the Heiberg algorithm relies on a careful selection of parameters. To a large extent, the choice of parameters represents partial knowledge about the nature of the problem, or about the environment in which segmentation is to take place. For example, for algorithms that are generalisable such that they can be used in more than one modality, object, or clinical problem, a common approach is to find a generally optimal set of parameters for each scenario, suitably defined on a test database through trial and error or by machine learning. In the case of the Heiberg algorithm and their implementation, provided freely online, a selection of pre-defined parameter sets is provided, each optimised for a particular *general* scenario: different types of MRI, segmentation of Left Ventricle vs Right Ventricle, segmentation from CT, etc. There are drawbacks to such 'scenario-based' approaches: Firstly, while a parameter set optimised on a training set adhering to a particular scenario serves as a good starting point, as we will demonstrate further on, it does not guarantee optimal results on particular images (even within the limits of the particular algorithm), or for particular setups and clinical contexts. Secondly, selecting an optimal set of parameters is normally a process which is largely intrinsic to the inner workings of an algorithm, offering little to no intuition on how they should be adjusted to accommodate changes in clinical context to ensure a more relevant outcome. Therefore, if a particular clinical environment has a slightly different setup to the one used for the algorithm training phase, and therefore has slightly different parameter requirements for optimal results (within the limits of the algorithm) than the ones provided by the manufacturer, tweaking that default parameter set to adjust it for their own setup is usually beyond the abilities of the clinician, because it does not translate to relatable clinical information. Therefore the clinician is more likely to simply accept the suboptimal parameter set (and by extension, a suboptimal segmentation result) *as is*, and simply try to take this into account clinically when weighing up the information. The main motivation behind our approach, therefore, is to enable the clinician to steer a segmentation algorithm towards results which are more relevant to their particular clinical context, by allowing them to introduce intuitive and clinically relatable information to the process; this could be performed once to adjust the the default parameter set to one more suitable to a particular clinical setup, or it could be performed on a per-case basis as required.

The intuition for our approach lies in the following key observation: Segmentation results which are 'better' — better, here, defined as results that are closer, in a mathematical sense, to the gold standard — will also produce estimates of physiological parameters — such as Ejection Fraction (EF) and Stroke Volume (SV) — which are 'better'. Our first premise, therefore, is derived by following this logic in reverse:

Premise 1: A segmentation result producing a large number of physiological parameters, which both individually and as a group are all 'better', is more likely to correspond to a 'better' segmentation. — Fig. 1 demonstrates this graphically.
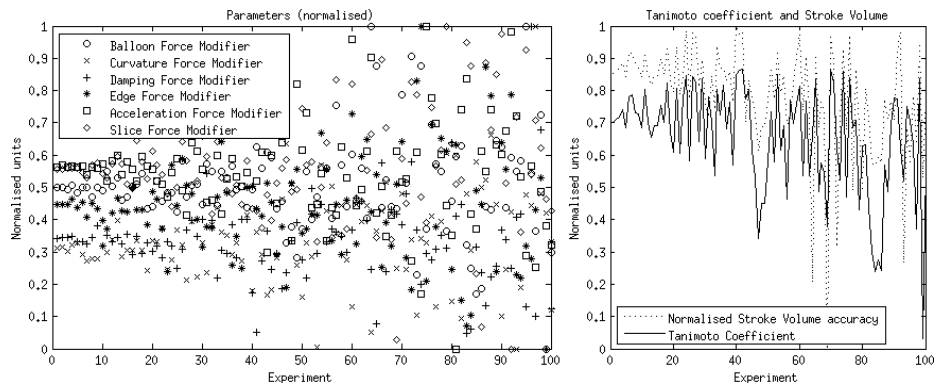
Fig. 1. Generating 100 segmentations using a default parameter set and increasing noise. Left: Distribution of algorithm parameters (normalised for comparison) for all experiments. Right: Comparison of physiological parameter accuracy and segmentation accuracy for all experiments; here stroke volume (SV) accuracy (A) is defined with respect to the Gold Standard (SV_G) as follows: $A = 1 - \frac{f(SV)}{f(SV)_{max}}$, where $f(SV) = |SV - SV_G|$ and $f(SV)_{max}$ is the maximum f value obtained within this set. There are three pertinent things to note: 1) As the parameters become increasingly noisy, accuracy tends to drop, but there are *still* occasions which produce good results. 2) The 'default' parameter set is not necessarily the best one; but it's difficult to predict a priori which one is. 3) Sets with even just *one* estimated physiological parameter being closer to the true value are visibly more likely to have higher accuracy; this is regardless of whether they originated from the less or more 'noisy' part of the experiment.

If we had the theoretical ability to explore all the possible values and combinations for each of the algorithm's parameters, we would obtain a set of segmentation results, covering all possible segmentation outcomes that are possible for a particular algorithm on a given image. We refer to this finite set, as the algorithm's *Segmentation Space*. Equivalently, the complete *Parameter Space* is the set consisting of all possible parameter sets, each mapping to a segmentation in the segmentation space.

While exploration of the full parameter space is generally infeasible, we can select samples from a focused region, which is most likely to correspond to more accurate segmentations. If we treat a set of N parameters as an N-dimensional vector, then a simple way of doing this is by selecting samples with an N-dimensional gaussian probability function centered at a point of interest. A reasonable choice for this would be the default parameter set/vector suggested by the algorithm itself. This hopefully should restrict the segmentation space to a subset of generally more accurate segmentations, which we could then fuse to obtain a fuzzy segmentation.

It follows from Premise 1, that if we introduce a bias in the fusion process, to favour segmentations that are 'better', the fused result should logically be biased towards a 'better' fused result.

Premise 2: In the presence of a segmentation subspace, biasing segmentation fusion towards results associated with better physiological parameters, should result in a better fused result overall, compared to an unbiased fusion

The practical implication of applying the above insights to any algorithm, is that we shift the focus from having to optimise highly unintuitive parameters intrinsic to the segmentation algorithm, to something that is more intuitive within the context of the task at hand – i.e. the physiological parameters – and which is therefore easier, and more relevant to non image-analysis specialists.

4   Methods

The Segment cardiac segmentation suite by Heiberg et al. (`http://segment.heiberg.se`) [4] was used to obtain left ventricle (LV) segmentations from a set of images, kindly provided by the University of Oxford Centre for Clinical Magnetic Resonance Research at the John Radcliffe Hospital, Oxford. This set was produced on a 3.0T Siemens Tim Trio whole-body MRI scanner using a 4D (i.e. 3D+time) TrueFISP cineMRI protocol, from a patient undergoing a post-PCA investigation, following a diagnosis of an Inferior MI; the set was anonymised appropriately and no other clinical or radiological details were available. The image set consisted of 25 timeframes of 8 Short-Axis (SA) slices at a resolution of 256×176 voxels, of size $1.5625 \times 1.5625 \times 8mm$. Manual segmentations of the left ventricle were provided by an expert clinician, which were used as a gold standard; this was obtained as per-slice 2D contours, drawn at 4×4 subresolution accuracy per in-slice image voxel, using the CMR42 cardiac imaging suite [5]. Full diastole was identified in timeframe 1, and full systole at timeframe 10. Data was processed using Matlab [6]; images were extracted from the DICOM files using a modified version of Laszlo Balkay's DICOM reader [7]; all other processing (including extraction of contours from CMR42 files) is the work of the authors.

A set of 100 segmentations was obtained by applying normally distributed random noise of linearly increasing standard deviation, on each of the default parameters (i.e. force modifiers; see Fig.1) provided by Segment for the case of SSFP MRI; the noise was generated with mean $\mu = initial\ modifier\ value$ for each parameter, and standard deviation σ taking values linearly from 0 to μ over the 100 experiments. Experiment 70, which was the best outcome in this set was retained as a reference to the best segmentation obtainable with this algorithm for this particular image. For each of the resulting segmentations, the following physiological parameters were derived:

- Volumes in systole (V_s) and diastole (V_d), defined as the number of voxels in the set of LV-labeled voxels in systole (LV_s) and diastole (LV_d) respectively
- Stroke Volume (SV) = $V_d - V_s$
- Ejection Fraction (EF) = SV/V_d
- Centre of mass (systole): A 3D-coordinate vector $C_s = [\bar{x}_i, \bar{y}_i, \bar{z}_i]^T, \forall\ i \in LV_s$
- Centre of mass (diastole): $C_d = [\bar{x}_i, \bar{y}_i, \bar{z}_i]^T, \forall\ i \in LV_d$
- Combined centre of mass: $C = (C_d + C_s)/2$

The weight each segmentation carries within the fusion process is determined by a measure of how close each of their physiological parameters is to a reference value; in particular, a suitable range around this reference value acts as a fuzzy constraint, that prevents bad segmentations, from a physiological-estimates point of view, from exerting much influence on the fused end-result. In practice, such values might be already available clinically (i.e. from a previous echocardiogram), or from known values. However, for the purposes of this paper, three different types of constraints were generated:

- Reference range derived from the 'default' segmentation (i.e. the segmentation resulting from the 'default' parameter set), using the median and inter-quartile range to define lower (l) central (c) and upper (u) reference values: This should produce the fuzzy analogue closest to the default case.
- Reference values derived from a very quick and crude initialisation process, where the user draws rough squares outside and inside the blood pool; thereby defining lower (l) and upper (u) constraint values for the reference range, with their average representing the central reference value (c).
- Reference values derived from the known Gold Standard. This should produce the best outcome which is possible from the algorithm, with respect to the known gold standard. Lower (l) and upper (u) constraint values for this case were set as $\pm$ 10% of the central value (c) for all physiological parameters, except for the distance from the centroid, which was set at the range of 0–10 voxels apart.

Weights were then calculated for each of n segmentations (S_n) from these reference values, by evaluating a fuzzy membership function on each of the estimated physiological parameters. We found that a good membership function was a gaussian membership function, with mean $\mu = c$ and standard deviation $\sigma = (u - l)/2$. A total weight (w) was then obtained by fuzzy conjunction of all the weights; this was evaluated separately for two Triangular-Norms[8]: Product (involving multiplication of all terms), and Gödel (involving selecting the minimum of the set as the weight, i.e. the "weakest link"). Segmentations were then fused by a simple weighted averaging process: $S_{fuzzy} = \sum_1^n w_n S_n$, and the fused result thresholded at 0.5; the resulting binary segmentation mask S was validated against the Ground Truth G using the Tanimoto Coefficient: $\frac{G \cap S}{G \cup S}$.

5   Results and discussion

The accuracy of the different segmentations is shown in Table 1. For comparison, the Tanimoto coefficient of the original algorithm with default parameters was 0.7016; the best possible outcome for the algorithm yielded a Tanimoto coefficient of 0.8671. While the Fuzzy equivalent of the default parameter set seems to be a bit lower for both Product and Gödel cases, it is very close (and indeed this is also the case visually; see Fig. 2), and it is in fact a better fuzzy equivalent than the simple averaging of all 100 segmentations without weighting, which
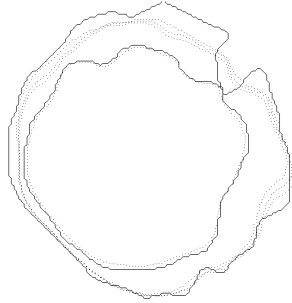
Fig. 2. Segmentation results on representative slice (5th): Continuous lines: inside – original default segmentation; outside – manual gold standard. Dotted lines, from innermost to outermost: a) result from physiological constraints from default case; b) result from crude manual initialisation c) result from initialisation from gold standard d) best segmentation from initial set.

	Default	Manual	Gold Standard
Product fuzzy logic	0.6745	0.7953	0.8183
Gödel fuzzy logic	0.6768	0.7536	0.8102

Table 1. Tanimoto coefficient of resulting fuzzy segmentations at diastole, after thresholding at 0.5 to obtain a binary result, and as compared against the gold standard (i.e. manual segmentations), for Product and Gödel fuzzy logic.

resulted in a Tanimoto coefficient of 0.6507. However, with respect to the best outcome, both the default case and its fuzzy analogues are poor by comparison.

There is clear improvement when more appropriate physiological parameters are provided as constraints. In the case of parameters derived from the Gold Standard (which would be equivalent, for instance to having those parameters provided by the clinician, e.g. via a different investigation or from prior knowledge), this comes very close to the 'best' result of the set. Furthermore, the rough manual initialisation is not far behind in terms of accuracy. In other words, even in the abscence of perfect physiological parameters, a rough estimate can still lead to a markedly better result. It is worth pointing out that it was only possible to identify the 'best' result via validation against the gold standard; therefore, in the absence of a gold standard, it would be very difficult to confidently identify the optimal parameter set. Our results demonstrate that by using the more intuitively generated physiological constraints in this fashion, we can achieve similarly good results as the best possible segmentation obtained through an optimal parameter set.

Figure 2 demonstrates the resulting contours, and shows the effect of our approach visually. Rather unsurprisingly, we note that the algorithm seems to retain its shape properties; in other words, if all segmentations in the set share common shape characteristics, the fused result is unlikely to produce a result which is structurally very different than the best result in the set. However, since the resulting surface is biased towards having similar physiological parameters as the gold standard, the final outcome should favour surfaces that are generally closer to it.

Conclusion

We have demonstrated a framework for producing a fuzzy equivalent segmentation from an existing algorithm, by exploring its parameter space to produce a segmentation space. This can then be fused in a weighted scheme, constrained by physiological parameters which can either be introduced by a clinician, much more intuitively than intrinsic algorithm parameters, or can be approximated by rough initialisation. The concept and framework can be generalised to any algorithm, such that instead of focusing on optimising intrinsic parameter sets for general cases, one would only need to explore the parameter space appropriately, and provide appropriate physiological constraints, which can be more intuitively defined, to produce better segmentations. The framework is particularly suited for medical images where the object in question has particular physiological properties that can then be represented via a fuzzy membership function and incorporated as a constraint; heart segmentation lends itself naturally to this problem, as it offers both physiological and anatomical constraints. Further work could focus on automating initialisation further, such as by using Haar features; introducing further types of physiological constraints, such as correctness of anatomical position based on other landmarks (e.g. defined as fuzzy spatial relationships of being "below the lung", "above the diaphragm" etc); and improving efficiency through parallelisation or a convergent approach to the acquisition of segmentation weights.

Acknowledgements

References

1. Petitjean, C. & Dacher, J.-N. "A review of segmentation methods in short axis cardiac MR images." Med Im Anal. 2011, 15, 169-184
2. R. E. Schapire, "The boosting approach to machine learning: An overview." Lecture notes in Statistics - New York - Springer Verlag, pp. 149172, 2003.
3. S. K. Warfield, K. H. Zou, & W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation." Medical Imaging, IEEE Transactions on, vol. 23, no. 7, pp. 903921, 2004.
4. Heiberg, E., Wigstrom, L., Carlsson, M., Bolger, A. F., & Karlsson, M. (2005, September). "Time resolved three-dimensional automated segmentation of the left ventricle." In Computers in Cardiology, 2005 (pp. 599-602). IEEE.
5. cmr^{42}. [software]. Circle Cardiovascular Imaging Inc., Calgary, Canada
6. MATLAB, v8.2 (R2013b). Natick, Massachusetts: The MathWorks Inc., 2012.
7. L. Balkay. DICOMDIR reader. University of Debrecen. 2011. Available online: http://www.mathworks.co.uk/matlabcentral/fileexchange/7926-dicomdir-reader
8. Klement, Erich Peter; Mesiar, Radko; & Pap, Endre. "Triangular Norms." Kluwer, Dordrecht, 2000