Shape and appearance priors for level set-based left ventricle segmentation

Ronghua Yang\textsuperscript{1}, Majid Mirmehdi\textsuperscript{1}, Xianghua Xie\textsuperscript{2}, David Hall\textsuperscript{3}

\textsuperscript{1}Department of Computer Science, University of Bristol, Bristol BS8 1TH, UK  
\textsuperscript{2}Department of Computer Science, University of Swansea, Swansea SA2 8PP, UK  
\textsuperscript{3}Department of Medical Physics and Bioengineering, University Hospitals Bristol NHS Foundation Trust, UK  
E-mail: majid@cs.bris.ac.uk

Abstract: The authors propose a novel spatiotemporal constraint based on shape and appearance and combine it with a level-set deformable model for left ventricle (LV) segmentation in four-dimensional gated cardiac SPECT, particularly in the presence of perfusion defects. The model incorporates appearance and shape information into a ‘soft-to-hard’ probabilistic constraint, and utilises spatiotemporal regularisation via a maximum a posteriori framework. This constraint force allows more flexibility than the rigid forces of shape constraint-only schemes, as well as other state of the art joint shape and appearance constraints. The combined model can hypothesise defective LV borders based on prior knowledge. The authors present comparative results to illustrate the improvement gain. A brief defect detection example is finally presented as an application of the proposed method.

\section{Introduction}

Single-photon emission computed-tomography (SPECT) is a commonly used approach for clinical evaluation of coronary artery disease. It visualises functional information of the left ventricle (LV) in three-dimension (3D), such as discontinuous perfusion scintigrams because of scarred or ischaemic myocardium. Gated cardiac SPECT also provides cardiac motion information along the time dimension.

The segmentation of LV epicardial and endocardial borders allows quantitative analysis of perfusion defects and cardiac function, and hence is of significant importance for diagnosis and clinical study. In SPECT, normal perfusion has high brightness in good contrast against a dark background, whereas perfusion defects have low or no brightness and hence low or no contrast against the background. Moreover, SPECT imaging can suffer from motion artefacts because of long imaging times \cite{1}, which is not a trivial task to compensate. As a result, accurate determination of the LV borders around the defect is difficult as there are few or no defined image features, such as gradients. In addition, the image data very often contains hot structures in the proximity of the LV, such as the liver, considered as perfusion artefacts that impose similar difficulties for correct LV segmentation. Furthermore, although resolution recovery techniques, such as \cite{2}, can be applied to increase SPECT spatial resolution, SPECT images have far lower resolution, as well as exhibit fewer anatomical landmarks (since the images suffer degradation because of various types of attenuation), in comparison with other modalities, such as Computed tomography (CT) and magnetic resonance imaging (MRI). Methods that are purely data-driven may only work well on SPECT images that do contain those image artefacts and ambiguities caused by perfusion at surrounding tissues, for example, paper \cite{3} which uses a conventional edge based active contour method. Prior knowledge of the LV shape is hence usually required to estimate myocardium borders and defective borders in particular.

Beyond such analysis in static LV data, segmentation of LV in 4D allows analysis of cardiac motion. However, because the low-resolution nature of SPECT imaging produces a partial volume effect, the images at end-systole tend to look brighter than end-diastole – known as the artefact of myocardial brightening \cite{4}. Hence, prior knowledge of cardiac motion is also required to estimate LV borders to guarantee correct motion estimation.

We propose a level-set framework that incorporates a novel shape and appearance-based spatiotemporal constraint into a general data-driven deformable model. The proposed constraint makes use of a priori knowledge in shape, appearance and their correlation, as well as spatiotemporal variations of the object under study, allowing more flexibility and reliability in the overall model. This helps cope with misleading image data such as defects, artefacts and temporal inconsistencies in gated cardiac SPECT.

In the next section, we review previous work on LV segmentation in cardiac SPECT data and other related issues. Section 3 covers charged active contour based on electrostatics (CACE), the basis for the proposed constrained CACE (CCACE) model. Section 4 gives details of CCACE, followed by experimental results of the model in Section 5. Finally, the paper is concluded in Section 6.
2 Previous work

A variety of deformable models have been proposed for LV segmentation in cardiac SPECT (e.g. [5–10]). Although bottom-up methods, for example [9] based on piecewise constant assumption, have been used, they generally suffer from image ambiguities, for instance, caused by under-perfusion. Top-down approaches in comparison are probably more robust, owning to the rigidity of the pre-defined models. For example, in [10], a predefined LV mesh is manually initialised and subsequently deformed to fit into local image gradients. Defective borders in SPECT can be hypothesised with these models as they sustain the shape of an LV during segmentation. However, these models have neglected the time dimension and deal with static LV only, missing the advantages that time consistent segmentation can offer [11]. Others have pursued a sequential segmentation approach, for example [12, 13], by taking the segmentation result of the previous timeframe as an initial guess for the current timeframe, but this generally leads to an underestimation of cardiac motion [14]. Paying more attention to the time dimension, [15] built a statistical model on the motion fields between the images at end-systole and end-diastole for better estimate of cardiac motion. Nonetheless, this approach, as well as other motion tracking algorithms [15, 16–18], track cardiac motion along sampled points in the image, which may not provide accurate segmentation of the LV shape.

There are also some elaborate spatiotemporal methods for better estimates of LV shape and motion on gated cardiac SPECT. Debureu et al. [11] and Charnoz et al. [19] implemented a geometric active contour to determine LV borders taking into account all the timeframes of the sequence to cope with background noise and provide more temporally smooth results. Montagnot and Delingette [14] used a pre-determined 4D shape model registered onto the input image sequence under shape and temporal constraints. Similar to [20–22], Kohlberger et al. [23] built a statistical constraint through principal component analysis (PCA) on training shapes. The constraint is applied to a level-set model searching LV borders in a Mumford–Shah’s formulation [24, 19]. We hereafter refer to this model as SCMS, that is, the Spatiotemporally Constrained Mumford–Shah’s model. The model in [25] was extended from [23] and inferred the constraint by utilising a kernel density estimator as suggested in [26] instead of the multivariate Gaussian model.

The main limitations of the above models would become apparent if they dealt with pathological cases. For example, the authors of [11, 19] assumed that the LV has constant pixel values, Montagnot and Delingette [14] relied on its own rigidity to hypothesise on the defective borders, and the constraints in [23, 25] were applied to the models rigidly, that is, with equal strength everywhere. Also, Kohlberger et al. [23] only allows local variations that the statistical priors can recover. Such shortcomings can compromise the accuracy of the estimate, especially on LVs with local variations and medium-to-severe perfusion defects. Furthermore, the constraints of these models are shape-based, with no mechanism proposed to incorporate appearance information correlating to the shape in the constraint forces. Since the shape of a perfusion does not always correspond to the LV shape in the presence of defects, we argue that these shape-based constraints still lack flexibility to deal with abnormal perfusion and hence can compromise segmentation accuracy. Appearance information, in addition to shape, should be considered.

In [27], Roussos and Paragios proposed a constraint via finding maximum probability density function (pdf) of the shape and combined it with a level-set model for object detection in optical images with noisy or missing data. This Self-Constrained Geodesic Active Regions model is hereafter referred to as SCGAR. Their probabilistic approach took into account shape variability, producing a soft-to-hard force, which allowed a wider range of shape variations than the models in [23, 25]. We will use this methodology as part of our approach towards LV detection in cardiac SPECT as LV shapes contain large degrees of irregular inter-class variations [28].

For more reliable application in SPECT, we take inspiration from the soft-to-hard probabilistic framework of SCGAR and replace its shape-based pdf objective function with a posterior that considers the shape and appearance of the whole sequence, leading to a novel shape and appearance-based spatiotemporal constraint derived with maximum a posteriori (MAP) estimation. Instead of globally aligning shapes when evaluating the pdf in SCGAR, we apply a global-to-local alignment. Meanwhile, as in [23, 25], PCA is employed in our model to reduce data dimension and statistically depict shape variations. However, unlike [23, 25], instead of applying PCA to the whole level-set domain, we apply it to the deformation fields of the level sets, characterised by transformation parameters, to further reduce data space and allow better control over temporal shape variations. Compared with current works on joint shape and appearance constraints such as [29–33], our soft-to-hard constraint allows more flexibility than their rigid forces. In addition, we argue that the probabilistic framework makes better use of the correlation between shape and appearance than those that linearly combine the two, for example [29–33].

In all, we present a model combining our recently proposed general-purpose deformable model for boundary detection, namely CACE [34], with a shape and appearance-based constraint that is capable of incorporating prior knowledge of correlated LV shape and appearance, as well as spatiotemporal variations. This is designed to cope with the irregularities in variations of LV shapes and motions and the extremely fuzzy gradients because of (temporal) perfusion defects and artefacts in cardiac SPECT. We refer to this combined model as CCACE.

3 Review of the CACE model

We now briefly review the CACE model [34] which is the foundation stone for the proposed CCACE method. CACE is a general-purpose active contour model that incorporates particle-based electrostatic interactions into the geometric active contour framework. It can be used for object boundary detection in a variety of images and applications. Embedded in level sets, CACE propagates under the influence of a bi-directional force field that simulates the electrostatic interaction between an image-derived point charge field and a charged contour. The force field leads the contour towards object boundaries and dynamically adapts as the contour evolves. It is a joint electrostatic force consisting of a boundary attraction force $F_A$ and a competition force $F_C$:

$$J = \lambda_A F_A + (1 - \lambda_A) F_C$$

(1)
\( F_\lambda = p_l \sum_{i=1}^{N_c} q_i \frac{s_i - r_l}{|s_i - r_l|^3}, \quad s_i \in X \)  
\( F_C = g(s_j)p_j \sum_{k=1,k \neq j}^{M} f(s_k)p_k \frac{s_j - s_k}{|s_j - s_k|^3}, \quad s_j \in X \)

\( \lambda_\alpha \) is a constant balancing the contributions from the two forces. \( f(\cdot) \) is the edge map of the input image, and \( g(s_j) \) is an edge stopping function (commonly defined as \( g = 1/(1 + f) \)). \( q_i \) is the negative fixed charge assigned to location \( r_l \), with magnitude proportional to its edge strength, that is, \( q_i = -f(r_l) \). \( p_j \) (or \( p_k \)) is the positive charge at location \( s_j \) (or \( s_k \)) on the active contour, and is set to unity for all \( s_j \), although \( p_j \) can be treated as a variable that can be tuned for specific applications. \( N_c \) and \( M \) are the number of negative and positive charges, respectively. \( \varepsilon_0 \) is the permittivity of free space and \( X \) is the set of all possible locations in image domain.

The attraction force \( F_\lambda \) acts as a vector field that leads each contour point towards the boundaries from both sides. The competition force \( F_C \): upon a contour is proportional to the inverted strength of the edge (i.e. \( g(s_j) \)) covered by this contour. Contours that are on the object boundaries endow most to the competition force with contributions proportional to edge strength. In other words, the competition force exerts most influence once any part of the snake reaches a boundary and repels ‘free’ contours nearby from reaching the already occupied boundary, forcing them to search for unoccupied boundaries. This repelling force is also designed in a way such that only contours in homogeneous regions will continue to deform according to both attraction and competition forces.

As the joint force field dynamically adapts by updating itself when a contour reaches a boundary, it brings flexibility in initialisation and better curve propagation towards object boundaries. The level-set \( \phi \) evolution of CACE is

\[ \frac{\partial \phi}{\partial t} = \lambda_\kappa \nabla \phi - (1 - \lambda_\kappa) J \cdot \nabla \phi \]

where \( \lambda_\kappa \) is a real positive constant and \( \kappa \) denotes the curvature of the contour (which regularises it). For more details of CACE the reader is referred to [34]. As mentioned earlier, we build CCACE on the CACE framework and effectively CCACE is CACE with application of specific constraints for 4D LV segmentation.

4 Proposed CCACE model

4.1 Overview

As shown in Fig. 1, the proposed model comprises of two stages: training and segmentation. Through the training stage, sets of pixel-wise Gaussian and spatiotemporal priors are obtained. The Gaussian priors include a prior image, a prior shape, an image variation term, a shape variation term and the correlation between image and shape. The spatiotemporal priors are the products of applying PCA to global-to-local transformation parameters of shape variations, and include mean of the parameters, eigenmodes (i.e. modes of variations), weights for the eigenmodes, as well as the covariance of the weights.

During segmentation, an initial surface embedded in level sets is placed in the input image. A constraint force is derived by finding the maximum of the multivariate pdf of the input image and shape (the evolving level sets) based on the Gaussian priors from the training stage. As the level sets evolve, the pdf continuously updates by aligning the prior image and shape with the input image and shape via global-to-local transformations to enable meaningful evaluation of the pdf. The alignment is regularised by prior distribution over a set of ‘spatiotemporal parameters’ of the transformations. The force derived from this regularised multivariate pdf, effectively a posterior, is a shape and appearance-based spatiotemporal constraint obtained via MAP estimation. The evolving level sets update according to the combination of CACE forces and this constraint to reach (and hypothesise missing or defective) LV borders. Next, we discuss the training and segmentation stages in detail.

4.2 Training

A training set \( \mathcal{M} = \{ (\bar{x}_i, \bar{\phi}_i) : i \in [1, N] \} \) of \( N \)-gated SPECT samples is constructed from manually labelled 4D sequences. Each sample \( (\bar{x}_i, \bar{\phi}_i) \) comprises of a pair – an image sequence \( \bar{x}_i \) and its corresponding LV shape sequence \( \bar{\phi}_i \), where \( \bar{x}_i \) is the \( i \)th training image sequence, and \( \bar{\phi}_i \) is the \( i \)th training shape sequence induced in level sets constructed by manual labelling on \( \bar{x}_i \). Each training sequence consists of \( K \) timeframes, that is, \( \bar{x}_i = \{ \bar{x}_{i,k} \in [1, K] \} \) and \( \bar{\phi}_i = \{ \bar{\phi}_{i,k} \in [1, K] \} \). Any known defect-free sequence can be selected as the reference shape sequence \( \phi = \bar{\phi}_{i,k} \in [1, K] \). To reduce non-linearity to fit in a Gaussian prior, each training shape \( \bar{\phi}_{i,k} \) is globally and locally aligned towards the corresponding timeframe of the reference shape sequence, that is, \( \bar{\phi}_i \), via a global-to-local registration method, whose details are given in the Appendix. The transformations recovered from these shape registrations are then applied to the training images accordingly to sustain the correspondence between training images and shapes. From the aligned training set, we derive two sets of priors: pixel-wise Gaussian priors and spatiotemporal priors.

Gaussian priors – We compute pixel-wise means of the aligned images \( \bar{I} = \{ \bar{I}_{i,k} \in [1, K] \} \), image standard deviations \( \bar{\sigma}_{i,k} \), means of the aligned shapes \( \bar{\phi} = \{ \bar{\phi}_{i,k} \in [1, K] \} \), shape standard deviations \( \bar{\sigma}_{\phi_{i,k}} \) and the correlation coefficients between image and shape \( \bar{\rho} = \{ \bar{\rho}_{i,k} \in [1, K] \} \). The bias \( \bar{I} \) and \( \bar{\phi} \) are also referred to as the prior image/appearance and prior shape sequence, respectively. As the training shapes are locally aligned, there is risk of bias in the priors towards the reference \( \bar{\phi} \), whose impact on model performance is however insignificant, as will be mentioned later at the segmentation stage. In practice, we choose a sample with a typical LV shape that has the least local variations from the mean shape of the training set, in order to further reduce the bias. Along with their standard deviations, the mean
image and shape sequences of our training set, for $K = 8$ and $N = 15$, are shown in Fig. 2.

**Spatiotemporal priors** – We seek parameterised temporal correlation among the spatial variations by applying two levels of PCA to a set of transformation parameters that describe how the training shape sequences $\phi_i$ vary from the prior shape sequence $\phi$. These transformation parameters, denoted as $\{\hat{\Theta}_{i,k}: (i, k) \in [1, N] \times [1, K]\}$, are recovered by mapping the prior shape sequence $\phi$ into the training shape sequences $\phi_i$ for $i \in [1, N]$, again via the global-to-local registration model (25). Note that in this mapping $\phi$ is the source shape and $\phi_i$ are the target shapes, whereas in the alignment of the training set above, $\phi_i$ is the source and $\phi$ the target. The reason for mapping here is that the distribution of $\hat{\Theta}_{i,k}$ can be used to regularise segmentation later during which the prior appearance and shape are aligned towards the input appearance and shape when evaluating the posterior.

The first-level PCA is applied to the recovered transformation parameters $\{\hat{\Theta}_{i,k}: i \in [1, N]\}$ at each timeframe $k$ individually, each giving a mean $\hat{\Theta}_k$, eigenmodes $P_k$ comprising the $m_k$ most significant modes of variations, and weights for the eigenmodes $\{b_{i,k}: i \in [1, N]\}$. The transformation parameters $\hat{\Theta}_{i,k}$ can then be approximated as

$$\hat{\Theta}_{i,k} = \hat{\Theta}_k + P_k b_{i,k} \quad (5)$$

This PCA captures the variations among shape transformations. We also consider that there is potential correlation among the global and local transformations which are both included in the PCA.

A further PCA is then applied to the weights of all timeframes, that is, $\{b_{i,k}: (i, k) \in [1, N] \times [1, K]\}$, and produces second-level eigenmodes $Q = [Q_1 \ldots Q_K]$ representing the $n$ most significant modes of variations, and weights $\{\hat{c}_i: i \in [1, N]\}$, giving the following representation

$$b_{i,k} = Q_k \hat{c}_i \quad (6)$$

The transformation parameters $\hat{\Theta}_{i,k}$ can be therefore represented in $\hat{c}_i$ as

$$\hat{\Theta}_{i,k} = \hat{\Theta}_k + P_k Q_k \hat{c}_i \quad (7)$$

Different to the first-level PCA, in this second-level one the entire sequence is taken as one dataset as an observation, hence the weights $\hat{c}_i$ do not relate to time dimension as $b_{i,k}$ does, that is, $\hat{c}_i$ is not indexed by $k$.

The weights from the first PCA statistically depict shape variations of every timeframe, and the second PCA on them produces parameters that control their temporal variations. This is similar to Cootes et al. [35] where an appearance model is built by applying second-level PCA to the results of a first-level PCA on shape and image greylevel data, respectively. These two levels of PCA give an approximation for a new set of transformation parameters

$$\Theta_k = \hat{\Theta}_k + P_k Q_k \hat{c} \quad (8)$$

where $\hat{c}$ is a varying parameter to be determined with many fewer dimensions than the transformation parameters $\Theta_k$. As it indirectly controls both spatial and temporal variations of the sequence by affecting its transformations, we refer to $\hat{c}$ as the 'spatiotemporal parameters'. Assuming normal distribution, the prior distribution of $\hat{c}$ is

$$P_{\hat{c}}(c) = \left(\frac{\sqrt{2\pi}}{|U|}\right)^{-1} \exp\left(-\frac{1}{2} c^T U^{-1} c\right) \quad (9)$$

where $U$ is the covariance matrix whose diagonal comprises the eigenvalues obtained from the two-level PCA.
To show the effect of varying the spatiotemporal parameters, we reconstruct transformations by varying each of the first two spatiotemporal parameters in turn, keeping the others zero, and apply the transformations to the prior shape, as shown in Fig. 3. It can be seen that the first parameter of \( c \) has obvious influence over global motions such as scaling and rotation of the LV, as well as local features such as the height of the septum. The second parameter, on the other hand, affects the LV shapes more subtly, for example, in changing the shape of the apical region and the ratio of the apical region over the whole LV size.

In summary, from the training stage, we obtain sets of Gaussian and spatiotemporal priors. The priors that are later used in the segmentation stage are listed in Table 1.

### 4.3 Shape and appearance-based spatiotemporal constraint

We now outline the derivation of our proposed constraint force based on the priors obtained during training and then give the combined formulation of CCACE in the next section. We use a variational framework to derive the constraint force that helps evolve the level sets to optimise a posterior, defined as the product of a shape and appearance-based multivariate pdf and prior distribution over the spatiotemporal parameters. The posterior measures the probability of an input shape and appearance to occur in a multivariate normal distribution characterised by the Gaussian priors obtained from the training stage, under spatiotemporal regularisation through prior distribution.

Given the input image sequence \( \mathbf{I} = \{ \mathbf{I}_k : k \in [1, K] \} \) and the evolving level sets \( \phi = \phi_k : k \in [1, K] \) placed on the image, the posterior is defined as

\[
P_A(\phi, c) = \prod_{k=1}^{K} \left( 2\pi \frac{1}{\Sigma_k} \right)^{-1} \exp \left( -\frac{1}{2} a_k^T \Sigma_k^{-1} a_k \right) \]

(10)

where

\[
a_k = \begin{bmatrix} a_{\phi_k} \\ a_{c_k} \end{bmatrix} = \begin{bmatrix} (1 - \lambda_k)(\phi_k - \phi_k(T)) \\ \lambda_k (\mathbf{I}_k - \mathbf{I}(T)) \end{bmatrix}
\]

(11)

\[
\sum_{k} = \frac{1}{1 - \mathbf{P}_k(T)} \begin{bmatrix} \tilde{\sigma}_{\phi_k}^2(T) & \tilde{\sigma}_{c_k}^2(T) \\ \tilde{\sigma}_{\phi_k}^2(T) & \tilde{\sigma}_{c_k}^2(T) \end{bmatrix}
\]

(12)

and \( T \) is the simplified representation of a global-to-local transformation model \( T(\Theta_k, \chi) \), details of which are given in the Appendix. The joint priors \( (\mathbf{I}, \phi) \) are aligned with the input appearance and shape \( (\mathbf{I}, \phi) \) through \( T(\Theta_k, \chi) \), for \( k \in [1, K] \). The transformations act on the priors so that the derived constraint force applies straightforwardly to the level sets and can be directly combined with CACE. Note that the impact of potential bias in the priors towards the reference \( \phi \) is largely alleviated via the alignment here. Positive constant \( \lambda_k \) balances the contributions from image and shape. \( \Sigma_k \) is the covariance between image and shape. \( \sigma_k \), a component of \( \Theta_k \), is the scaling factor at \( k \)th timeframe. The reason for applying \( \sigma_k \) to \( \phi_k \) that level-set representation is not invariant to scaling – see [27] for details. As both \( \phi \) and \( c \) are unknown, a coupled optimisation task can optimise the posterior

\[
\langle \hat{\phi}, \hat{c} \rangle = \arg \min_{\phi, c} E_A(\phi, c)
\]

(13)

where \( E_A(\phi, c) \) is the posterior-based objective function defined as

\[
E_A(\phi, c) = -\int_X \log(P_A(\phi, c)) \, dx
\]

\[
= \sum_{k=1}^{K} \int_X \left( \log \left( \frac{1}{(2\pi)^{\frac{1}{2}} |\Sigma_k|} + \frac{1}{2} a_k^T \Sigma_k^{-1} a_k \right) \right) dx
\]

\[
+ \log \left( \sqrt{2\pi e |U|} \right) + \frac{1}{2} c^T U^{-1} c
\]

(14)

\( X \) represents all the possible locations in the shape domain [Note that the level-set shape domain coincides with the image domain.]. \( E_A(\phi, c) \) considers both shape and appearance information and is evaluated with variations
between the priors \((I, \phi)\) and the input terms \((I, \phi)\) taken into account. In particular, only spatiotemporally plausible variations are accounted for owing to regularisation that biases the transformation parameters towards their most probable values. This keeps the model within plausible shapes and motions. The level sets evolve towards the maximum of the posterior to give the best probabilistic estimate over the shape in the input image.

### 4.4 Level-set evolution of CCACE

Combining the constraint component with CACE, the motion equations for CCACE are given in (15) and (16), through which \(\phi\) and \(c\) update iteratively and simultaneously until equilibrium

\[
\frac{\partial \phi_k}{\partial t} = \lambda_c \left( \lambda_k \kappa |\nabla \phi_k| - (1 - \lambda_k) U \cdot \nabla \phi_k \right)_{\text{CACE}} + (1 - \lambda_c) \frac{s_k}{1 - \bar{\rho_k}(T)} \left[ -(1 - \lambda_j) \frac{s_j \phi_k - \bar{\phi}_k(T)}{\bar{\sigma}_j^2(T)} \right] \left( \frac{\partial \bar{\sigma}_j^2(T)}{\partial \phi_k(T)} \right)_{\text{Shape}} - \lambda_j \bar{\rho}_k(T) \frac{I_k - \bar{I}_k(T)}{\bar{\sigma}_j^2(T) \bar{\phi}_j(T)} \left( \frac{\partial \bar{\phi}_j(T)}{\partial \phi_k(T)} \right)_{\text{Appearance}}
\]

\[
\frac{\partial c}{\partial t} = -\left( \frac{\partial E_d(\phi, c)}{\partial \Theta} \right)' \left[ Q_{k}^t P_{p}^t \ldots Q_{k}^t P_{p}^t \right] - U^{-1} c \tag{16}
\]

where \(\Theta = \Theta_k; k \in [1, K]\). In (15), the first term represents the data-driven evolution of CACE, whereas the second is the constraint force, derived based on (13), with \(\lambda_c\) balancing the contributions from the two. The constraint force consists of a shape and an appearance term. \(c\) also updates according to (16) at each iteration of \(\phi\)'s evolution to keep the prior image and shape aligned with the input image and shape.

To compute the partial derivatives \((\partial E_d(\phi, c) / \partial \Theta)\), we denote the transformation parameter as \(\Theta_j = \theta_{k, j}; j \in [1, N_\Theta]\), where \(N_\Theta\) is the length of \(\Theta_k\), so that \(\theta_{k, j}\) refers to the \(j\)th component of \(\Theta_k\). Thus, \((\partial E_d(\phi, c) / \partial \Theta)\) can be formulated as in (17).

\[
\frac{\partial E_d(\phi, c)}{\partial \theta_{k, j}} = \int \left[ \frac{1}{\Sigma_k^2} \frac{\partial \Sigma_k}{\partial \theta_{k, j}} + \frac{1}{2} \frac{\partial \Sigma_k^{-1}}{\partial \theta_{k, j}} a_k + a_k \Sigma_k^{-1} \frac{\partial a}{\partial \theta_{k, j}} \right] dx
\]

\[
= \int \left[ \frac{\partial a_k}{\partial \theta_{k, j}} a_k \frac{\partial \bar{\sigma}_j^2(T) - \bar{\rho}_k(T) \bar{\phi}_j(T)}{(1 - \bar{\rho}_k(T)) \bar{\sigma}_j^2(T) \bar{\phi}_j(T)} \right] \left( \frac{\partial \bar{\phi}_j(T)}{\partial \theta_{k, j}} \right)_{\text{Shape}}
\]

Note that because of the presence of scaling factor in front of the level sets in (14), the partial derivatives of \(a_k\) w.r.t. \(s_k\) in (18) take a different form from other global and local transformation parameters in (19).

According to (15), the level sets evolve under the influence of the force field in CACE and the proposed constraint force. The shape term drives the level sets towards the aligned prior shapes. In the denominator, the squared deviation allows shape variability, hence the shape driving force is applied in a soft-to-hard manner. Meanwhile, the appearance term penalises or strengthens the constraint force depending on the correlation between image and shape. The more distant the input image from the aligned prior image, the stronger the term is, where image variability is also taken into account. It plays the role of an extra criterion for constraining the level sets and brings further flexibility to the constraint.

### 4.5 Initial conditions

Although CCACE (or CACE) is less sensitive to initial placement, it is still not entirely independent from the initial conditions. Hence, we devise an automated initialisation method for CCACE for higher efficiency, reliability and practicality of the model.

Given the fact that the closer the initial shape to the final shape to be recovered, the more efficient the segmentation process is, we aim to find a rough shape estimate over the
Towards this end, we set \( \lambda = 1 \) in (14), so that the objective function \( E_A(f, c) \mid l = 1 \) takes information of greylevel intensities only (with shape information ignored), and measures the probability of the input image to occur. Note that the correlation coefficients \( \rho_k \) can be assumed zero and \( \phi \) can be ignored for \( E_A(f, c) \mid l = 1 \). We then find the minimum of \( E_A(f, c) \mid l = 1 \), by updating the spatiotemporal parameters \( c \), starting from 0, according to a motion equation taking the same form as (16), except that the partial derivatives in it are rewritten as

\[
\frac{\partial E_A(f, c)}{\partial t_{k,j}} \bigg|_{l = 1} = \int \left[ \frac{\partial \sigma_f(T)}{\partial t_{k,j}} \left( 1 - \frac{a_{ik}^2}{\sigma_f(T)} \right) + \frac{\partial a_{ik}}{\partial t_{k,j}} \frac{a_{ik}}{\sigma_f(T)} \right] dx
\]

With the recovered spatiotemporal parameters, denoted as \( \hat{c} \), and subsequently the recovered transformation parameters \( \Theta \), a probabilistic image estimate for input image \( I_k \) can be obtained and denoted as \( \hat{I}_k(T; \Theta_k; x) \). Its corresponding shape estimate is therefore \( \phi_k(T; \Theta_k; x) \). We set \( \phi_k = \phi_k(T; \Theta_k; x) \) and \( c = \hat{c} \) as the initial conditions for the coupled evolution task in (15) and (16). \( \lambda \) is then set to 0.5 so that both shape and appearance are considered equally during the evolution.

## 5 Experimental results

We used 15 sets of gated cardiac SPECT sequences of healthy LVs with normal perfusion for training, and applied CCACE to another 14 sequences for performance evaluation, two of which are pathological cases and present perfusion defects. All patient studies were clinical studies acquired following standard protocols. The resolution of the data was \( 64 \times 64 \times \alpha \times 8 \) where \( \alpha \) varied between 25 and 58. Throughout

![Fig. 4  CACE and other models on mid-slices containing perfusion defects and artefacts (marked by white arrows)](image)

From left, each column: input slice, results of CPM, geodesic snake, GVF geodesic snake, CACE and ground truth

![Fig. 5  Iterations of SCMS (rows 1–3), SCGAR (rows 4–6) and CCACE (rows 7–9) on gated cardiac SPECT with artefacts and severe perfusion defects. Final row of each model shows the results overlaid on the ground truth](image)
the experiments, we set $\lambda_c = 0.4$ and $\lambda_I = 0.5$ which were determined empirically. For each $P_k$ and $Q$, the 10 most significant modes were chosen to explain 95% of the variations, that is, $m_k = 10$ for $k \in [1, K]$ and $n = 10$.

For comparison, we also implemented the SCMS and SCGAR models based on [23] and [27], respectively. Since no particular initialisations are specified for these two models, we initially set them as the prior shapes for each example. Note, unlike the 4D nature of SCMS and CCACE, SCGAR deals with each timeframe individually.

We use the following colour scheme in the results that follow: model results in red, ground truth in yellow and localised defect regions in cyan.

5.1 Subjective assessment

To illustrate the necessity of the use of priors, in Fig. 4 we show typical results from a number of data driven methods, including charged particle model (CPM) [36], geodesic active contour [37], GVF geodesic [38] and our base model, CACE. Although CACE has successfully recovered the shapes of the perfusion, it has not outlined the true LV borders according to the ground truth, as the perfusion state does not reflect the true shape of the LV in the presence of defects or artefacts. This is because CACE evolves towards any image gradients it can capture, without considering the anatomical features of the object it is dealing with. Any

![Fig. 6 SCMS (rows 1–4), SCGAR (rows 5–8) and CCACE (rows 9–12) on slices selected from the sequence](image-url)
other general data-driven method that only considers image features is likely to suffer the same problem, as demonstrated in the examples. Hence, the rest of the comparative analysis is focused on methods that are using prior-based segmentation.

Fig. 5 shows iterations of SCMS (rows 1–3), SCGAR (rows 4–6) and CCACE (rows 7–9) for an LV with severe myocardial infarction in the inferior wall (consistent perfusion defects present throughout the sequence as well as perfusion artefacts because of uptake in the liver). For each model snapshots are shown in the first two rows, whereas final results against ground truth are shown in the third row. For visual assessment, clearly the less yellow (or red) is seen in the third row for each case better the results. Fig. 6 shows the results in selected 2D slices.

Large deviations between the SCMS results arise because SCMS only allows global scaling and translation and local variations that have occurred among the training sets, thus is strictly restricted to only a pre-defined range of variations. Meanwhile, SCGAR is more flexible than SCMS, as it takes into account rotation in addition to scaling and translation, and allows the model to locally deform based on region forces. It can therefore cope with more variations than SCMS and achieves higher accuracy in many areas. However, the SCGAR model tends to merge over the defect areas as marked by white arrows in Fig. 6 (1st column of SCGAR results), because of its data-driven forces, whereas CCACE sustains a plausible shape and successfully differentiates the epicardial from the endocardial borders in these areas. SCGAR has large deviations from the borders in the lateral segments as marked by green arrows (3rd column of SCGAR results), whereas CCACE converges more closely around them. This shows CCACE has more flexibility as its constraint is more alert to defect regions and less sensitive to lateral segments.

The limitation of SCGAR is due to the fact that the model only removes global variations between the prior and the input shape, thus there remain large deviations between them in areas containing local variations such as the defect region and lateral segments. Hence, the ratio between the deviations over plausible variations (e.g. in lateral segments) and implausible variations (e.g. in defect areas), and the resulting constraint forces which are proportional to these deviations, tend to be small. In contrast, as plausible variations between the joint priors and the input shape and image are accounted for through the regularised global-to-local transformations in CCACE, the deviations between them are diminished over plausible variations and kept large over implausible ones. As a result, the magnitudes of the constraint forces over the defects are

Fig. 7 SCMS (rows 1–2), SCGAR (rows 3–4) and CCACE (rows 5–6) on a large LV with long apical region
Black arrows indicate significant areas of mismatch against ground truth
significantly stronger than the lateral segments and other normal regions. CCACE can therefore give a good estimate over the defective borders without much compromise of its accuracy in other regions.

In addition, the appearance term in CCACE makes crucial contributions to the constraint by preventing the model from merging over the defect regions, because there is significant difference between the intensities of the defect and the aligned prior image, leading to a large appearance term. Meanwhile, in normal perfusion areas the appearance term exerts little influence because of small intensity differences. Hence, this term plays the role of a meaningful criterion that brings more flexibility than shape-based constraints only.

Figs. 7 and 8 show the models’ results on an example of a large LV with a long apical region in particular. SCMS and SCGAR fail to produce accurate estimates especially in the apical regions as marked by black arrows, whereas CCACE has achieved much better results.

5.2 Quantitative evaluation

The Jaccard coefficient is a commonly used statistic for similarity (or diversity) measure between two sets. It is the ratio between the size of the intersection and the size of the union of two sets, \( A \) and \( B \), under comparison: 

\[
\frac{|A \cap B|}{|A \cup B|}
\]

We compute the Jaccard coefficient between the
CCACE results are highlighted in bold. Column managed to outline the LV shapes, but errors occur not sequence with artefacts, attenuation and large local variations. CCACE is 94.2% on a normal perfusion sequence with time dimension than the other two models, attributed to its only in the apical and basal but also in middle regions with large local variations.

5.3 Computation time

Using a 2.8 GHz Linux PC running uncompiled Matlab code, the computational times for the SCMS, SCGAR and CCACE models are given in Table 3. The training stage of CCACE takes significantly longer than the other two models, because of the extra local transformations involved and the fact that the global-to-local transformations are applied to each dataset twice (the first one for alignment and the second one for spatiotemporal

<table>
<thead>
<tr>
<th>D</th>
<th>M</th>
<th>Jaccard coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>Timeframes</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>53.1</td>
</tr>
<tr>
<td>B</td>
<td>55.6</td>
<td>47.8</td>
</tr>
<tr>
<td>C</td>
<td>94.2</td>
<td>92.9</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>59.2</td>
</tr>
<tr>
<td>B</td>
<td>63.2</td>
<td>63.0</td>
</tr>
<tr>
<td>C</td>
<td>93.5</td>
<td>92.8</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>63.8</td>
</tr>
<tr>
<td>B</td>
<td>62.1</td>
<td>62.6</td>
</tr>
<tr>
<td>C</td>
<td>93.5</td>
<td>92.0</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>61.2</td>
</tr>
<tr>
<td>B</td>
<td>65.2</td>
<td>63.9</td>
</tr>
<tr>
<td>C</td>
<td>93.3</td>
<td>92.2</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>58.4</td>
</tr>
<tr>
<td>B</td>
<td>57.7</td>
<td>53.8</td>
</tr>
<tr>
<td>C</td>
<td>93.2</td>
<td>93.9</td>
</tr>
<tr>
<td>6 (Fig. 7)</td>
<td>A</td>
<td>48.3</td>
</tr>
<tr>
<td>B</td>
<td>48.9</td>
<td>41.8</td>
</tr>
<tr>
<td>C</td>
<td>90.6</td>
<td>88.0</td>
</tr>
<tr>
<td>7* (Fig. 5)</td>
<td>A</td>
<td>63.0</td>
</tr>
<tr>
<td>B</td>
<td>62.5</td>
<td>57.9</td>
</tr>
<tr>
<td>C</td>
<td>87.3</td>
<td>85.4</td>
</tr>
<tr>
<td>8* (Fig. 5)</td>
<td>A</td>
<td>65.1</td>
</tr>
<tr>
<td>B</td>
<td>62.3</td>
<td>62.4</td>
</tr>
<tr>
<td>C</td>
<td>85.8</td>
<td>83.6</td>
</tr>
<tr>
<td>9</td>
<td>A</td>
<td>59.8</td>
</tr>
<tr>
<td>B</td>
<td>63.2</td>
<td>62.2</td>
</tr>
<tr>
<td>C</td>
<td>85.4</td>
<td>86.7</td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td>55.2</td>
</tr>
<tr>
<td>B</td>
<td>68.4</td>
<td>66.9</td>
</tr>
<tr>
<td>C</td>
<td>84.7</td>
<td>85.6</td>
</tr>
<tr>
<td>11</td>
<td>A</td>
<td>57.0</td>
</tr>
<tr>
<td>B</td>
<td>61.5</td>
<td>58.9</td>
</tr>
<tr>
<td>C</td>
<td>84.6</td>
<td>79.6</td>
</tr>
<tr>
<td>12</td>
<td>A</td>
<td>70.9</td>
</tr>
<tr>
<td>B</td>
<td>78.8</td>
<td>81.9</td>
</tr>
<tr>
<td>C</td>
<td>82.6</td>
<td>81.5</td>
</tr>
<tr>
<td>13</td>
<td>A</td>
<td>49.8</td>
</tr>
<tr>
<td>B</td>
<td>40.8</td>
<td>31.0</td>
</tr>
<tr>
<td>C</td>
<td>823</td>
<td>68.8</td>
</tr>
<tr>
<td>14</td>
<td>A</td>
<td>48.8</td>
</tr>
<tr>
<td>B</td>
<td>47.1</td>
<td>41.5</td>
</tr>
<tr>
<td>C</td>
<td>80.9</td>
<td>76.1</td>
</tr>
</tbody>
</table>

Column ‘D’: dataset indices. Column ‘M’: model names, where ‘A’, ‘B’ and ‘C’ represent SCMS, SCGAR and CCACE, respectively, and CCACE results are highlighted in bold. Column ‘O’: the overall accuracy for each sequence.

Table 3: Computational times for the models

<table>
<thead>
<tr>
<th>Models</th>
<th>SCMS</th>
<th>SCGAR</th>
<th>CCACE</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>training</td>
<td>segmentation</td>
<td>51 min</td>
</tr>
</tbody>
</table>

© The Institution of Engineering and Technology 2013

analysis). Each global-to-local transformation in CCACE training requires 54s on average, whereas the global transformation during SCMS and SCGAR training needs 25 s. The segmentation stage of CCACE takes more time than the other two models also because of the extra local transformations as well as the more complex computation of energy minimisation. In summary, CCACE clearly pays a price for greater accuracy.

5.4 Defect detection

In this section, we present a brief defect detection study to illustrate the potential of the proposed method. One of the ultimate goals of LV segmentation is to assist diagnosis of heart disease, e.g. finding myocardial infarction by localising perfusion defect. With the image correctly segmented, defect detection is fairly simple since the scope of detection is within the myocardium and the defects present much lower brightness than normal perfusion. We accomplish this detection task on the two example image sequences with defects) by running CACE on the segmented images obtained from CCACE’s results. Other active contours with robust handling of weak gradients could also have been used.

We first define the segmented image as \( H(\tilde{\phi}_k) I_k \) where \( \tilde{\phi}_k \) is the level-set LV shape recovered by CCACE and \( H(\cdot) \) is a Heaviside function. The joint electrostatic force field is then derived for the segmented image according to (1) to guide CACE evolution towards defect boundaries. To automate the detection, we use the divergence-based initialisation method proposed in [39], by finding divergent points of the initial force field and selecting those in the dark areas.
In this paper we have proposed the CCACE model which recovers LV on gated cardiac SPECT by probabilistically determining its boundaries based on image gradients under a shape and appearance-based spatiotemporal constraint within a MAP framework. CCACE a) makes use of global-to-local rather than global registration to account for a shape and appearance-based spatiotemporal constraint, recovering LV on gated cardiac SPECT by probabilistically recovering LV on gated cardiac SPECT images using deformable mesh modeling.

The model has shown promising results on gated cardiac SPECT even on data with artefacts or severe perfusion defects. Further improvements may be achieved, for example, by utilising the kernel density estimator as suggested in [26] towards better evaluation of prior distribution of spatiotemporal parameters, and by applying a more robust optimisation scheme such as the trust region method.

7 References
23 Kohlberger, T., Cremers, D., Rousson, M., Ramaraj, R., Funka-Lea, G.: ‘4D shape priors for a level set segmentation of the left myocardium in SPECT sequences’ (MICCAI, 2006), vol. 9, pp. 92–100
26 Rousson, M., Cremers, D.: ‘Efficient kernel density estimation of shape and intensity priors for level set segmentation’ (MICCAI, 2005), vol. 8, pp. 335–342
31 Huang, X., Li, Z., Metaxas, D.: ‘Learning coupled prior shape and appearance models for segmentation’ (MICCAI, 2004), pp. 60–69
The global-to-local registration method proposed for CCACE combines rigid transformation and incremental free form deformation (IFFD) [40] and applies a sum of squared differences (SSD) criterion with a smoothness constraint, defined as

$$E_f(\Psi; x) = (1 - \lambda_s) \int_x \left( s \phi_d - \phi_s(T(\Psi; x)) \right)^2$$

$$+ \lambda_s \left( \frac{\partial \delta T_i}{\partial x} \right)^2 + \left( \frac{\partial \delta T_i}{\partial y} \right)^2 + \left( \frac{\partial \delta T_i}{\partial z} \right)^2 \right) dx$$ (25)

where $\phi_s$ is the source shape, which is to be registered onto the target shape $\phi_d$. The constant $\lambda_s$ balances between the SSD measure and the smoothness constraint. The transformation model $T(\Psi; x)$ is defined as

$$T(\Psi; x) = s A_R (x + \delta T_i) + \begin{bmatrix} h_x \\ h_y \\ h_z \end{bmatrix}$$ (26)

where

$$A_R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \omega_z & \sin \omega_z \\ 0 & -\sin \omega_z & \cos \omega_z \end{bmatrix} \begin{bmatrix} \cos \omega_y & 0 & -\sin \omega_y \\ 0 & 1 & 0 \\ \sin \omega_y & 0 & \cos \omega_y \end{bmatrix}$$

$$\delta T_i = \sum_{a=0}^3 \sum_{b=0}^3 \sum_{c=0}^3 B_3(\alpha) B_4(\beta) B_5(\gamma) \left[ \begin{array}{c} \delta P^x_{i,j,k+m} \\ \delta P^y_{i,j,k+m} \\ \delta P^z_{i,j,k+m} \end{array} \right]$$ (27)

$\Psi$ is the global-to-local transformation parameters defined as $s, \omega_x, \omega_y, \omega_z, h_x, h_y, h_z, \delta P$. The rigid transformation is constituted of scaling $s$, rotation $(\omega_x, \omega_y, \omega_z)$ and translation $(h_x, h_y, h_z)$. $\delta P$ is the displacement of an $L_p \times M_p \times N_p$ mesh grid of control points overlaid on the domain with uniform spacings $\delta_x, \delta_y$ and $\delta_z$, and $\delta P = \{(\delta P^x_{i,m,n}, \delta P^y_{i,m,n}, \delta P^z_{i,m,n}) | (i, m, n) \in [1, L_p] \times [1, M_p] \times [1, N_p]\}$. The indices $i, j, k$ denote the position of the control point at the top left corner of the cell that contains $x$, and $x, y, z$ are the relative positions of $x, y, z$ within the cell, that is, $i = [x/\delta_x] - 1, j = [y/\delta_y] - 1, k = [z/\delta_z] - 1$ and $x = x/\delta_x - [x/\delta_x], y = y/\delta_y - [y/\delta_y], z = z/\delta_z - [z/\delta_z]$. $B_i(\alpha)$ is the $i$th basis function of the cubic B-spline

$$B_0(\alpha) = (1 - \alpha)^3/6$$ (29)

$$B_1(\alpha) = (3\alpha^3 - 6\alpha^2 + 4)/6$$ (30)

$$B_2(\alpha) = (-3\alpha^3 + 3\alpha^2 + 3\alpha + 1)/6$$ (31)

$$B_3(\alpha) = \alpha^3/6$$ (32)

$B_j(\beta)$ and $B_k(\gamma)$ are similarly defined.