



A robust and accurate center-frequency estimation (RACE) algorithm for improving motion estimation performance of SinMod on tagged cardiac MR images without known tagging parameters[☆]

Hong Liu^{a,b}, Jie Wang^{a,b}, Xiangyang Xu^{a,b,*}, Enmin Song^{a,b}, Qian Wang^{b,c}, Renchao Jin^{a,b}, Chih-Cheng Hung^d, Baowei Fei^e

^a School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

^b Key Laboratory of Education Ministry for Image Processing and Intelligence Control, Wuhan, Hubei 430074, China

^c Department of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan, Hubei 430073, China

^d School of Computing and Software Engineering, Southern Polytechnic State University, Marietta, GA 30060, USA

^e Quantitative Bioluminescence Laboratory, Emory University School of Medicine, Atlanta, GA 30322, USA

ARTICLE INFO

Article history:

Received 20 January 2014

Revised 30 April 2014

Accepted 24 July 2014

Keywords:

Tagged cardiac MR image

Motion estimation

SinMod

Center-frequency

Mean-shift

Two-direction-combination

ABSTRACT

A robust and accurate center-frequency (CF) estimation (RACE) algorithm for improving the performance of the local sine-wave modeling (SinMod) method, which is a good motion estimation method for tagged cardiac magnetic resonance (MR) images, is proposed in this study. The RACE algorithm can automatically, effectively and efficiently produce a very appropriate CF estimate for the SinMod method, under the circumstance that the specified tagging parameters are unknown, on account of the following two key techniques: (1) the well-known mean-shift algorithm, which can provide accurate and rapid CF estimation; and (2) an original two-direction-combination strategy, which can further enhance the accuracy and robustness of CF estimation. Some other available CF estimation algorithms are brought out for comparison. Several validation approaches that can work on the real data without ground truths are specially designed. Experimental results on human body in vivo cardiac data demonstrate the significance of accurate CF estimation for SinMod, and validate the effectiveness of RACE in facilitating the motion estimation performance of SinMod.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Compared with ultrasound and X-ray computerized tomography, magnetic resonance imaging (MRI) [1] has the advantage of excellent soft tissue contrast. With MRI, structures and functions of tissues can be evaluated [2]. However, in MRI, tracking motion of moving tissues is quite difficult. The MRI tagging (MRIT) technique [3,4] was developed to address this problem. MRIT produces distinguishable labels (the so-called tags, generally presented as regular patterns) and tags them regularly into the interested tissue. The tags then move synchronously with the tissue particles to which they attach, making

the motion of any particle be capable of being tracked. With the tracked motion, modeling, understanding, quantifying and further analyses of the tissue all become achievable [5].

The tagged cardiac magnetic resonance (TCMR) image has demonstrated its significant role in clinical diagnosis and treatment of the in vivo heart by offering sufficient information of cardiac motion [6,7]. So far, plenty of useful methods have been proposed for estimating/tracking motion from TCMR images. Some of them first detect and track the distinct tag features such as tag lines and tag intersections, and then interpolate the local myocardial motion at any pixel location. Some others directly obtain the dense (mostly at the pixel level) motion displacement field (MDF), which describes the local motion between two TCMR images at the specified locations [2,7]. Methods of the latter category are preferred since they usually run faster and achieve higher motion estimation accuracy. Among these methods, the harmonic-phase analysis (HARP) method [8–14] and the local sine-wave modeling (SinMod) method [15–19] are regarded as the two most outstanding ones.

The kernel principle of the HARP method is that, for each myocardial particle, the phase related to a tag-direction, which

[☆] This work is supported by the National Natural Science Foundation of China (Grant No. 61075010), National Natural Science Foundation of China (Grant No. 61370179), Fundamental Research Funds for the Central Universities of China (Grant No. 2014TS009), and National Key Technology Research and Development Program of China (Grant No. 2012BAI23B07).

* Corresponding author at: School of Computer Science and Technology, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, Hubei 430074, China. Tel.: +86 27 87792212.

E-mail address: xuxy@hust.edu.cn (X. Xu).

describes the relative position relation between that particle and its adjacent tags along that tag-direction, remains unchanged no matter to where the particle moves. For two-dimensional (2-D) TCMR images tagged in two crossed in-plane directions, each particle has two such phases. In this context, HARP performs by extracting two phase maps from each image, and searching for the pairs of two positions from two sequential images, respectively, where the two positions should satisfy the following two conditions: (1) their phases are correspondingly equal; and (2) they are spatially close to each other. Once two such positions are found, they are considered to match to the same myocardial particle. Thus, the motion of that particle between the two images can be accordingly calculated. Due to its full automaticity, high estimation accuracy and fast processing speed, HARP has become a state-of-the-art motion estimation method for TCMR images.

The SinMod method was developed based on HARP. The difference is that SinMod concentrates on the image pixel rather than the material particle. In the local environment of a pixel, SinMod models the intensity distribution as the summation of two moving local sine wave-fronts (in the case of 2-D TCMR images with two tag-directions), which associates the local motion along a specific in-plane direction at that pixel with the frequency and phase-shift of the local sine-wave propagating along that direction (the so-called wave-direction, which is perpendicular to the tag-direction it corresponds to). Following this thought, SinMod develops a particular frequency analysis approach to work out the local frequencies and local phase-shifts for all pixels, and makes use of them to generate a dense, pixel-level MDF estimate. According to Arts et al. [15], the experimental results on both simulated and real TCMR data demonstrate that SinMod runs as fast as HARP, and performs better with respect to accuracy of displacement detection, noise reduction, and avoidance of artifacts. In general, SinMod is believed to have a better comprehensive performance as compared with HARP.

In this paper, we focus on an important issue: seeking an effective approach to accurately estimate the center-frequency (CF), which plays a significant role in the SinMod method for constructing the essential frequency-domain band-pass filters and calculating the local frequencies. We present the concept of accurate CF estimation because as we notice, the motion estimation accuracy of SinMod is fairly dependent on the CF selection. A better selection of the CF could yield a more accurate motion estimate; on the contrary, a poor selection might result in large local estimation errors. However, to the best of our knowledge, we have not discovered any specific CF selection/estimation approach in related works of SinMod yet. The original study [15] only provided a general idea for how to determine the CF but without an unambiguous algorithm; the other works using SinMod [16–19] were usually concerned about the application and comparison, but not the detailed implementation.

As a simple solution, the CF can be selected artificially. However, this way surely breaks the automaticity of SinMod, and moreover, it can hardly guarantee the selection accuracy of CF. In fact, based on the principle of the SinMod method, as long as the tagging parameters including the initial tag spacings and tag-direction angles set by the applied MRIT approach are all known, the ground truth of CF can be immediately computed. But it is more often the case that we only have the image data itself. To accurately select the CF under the circumstance that those mentioned parameters are unknown, we specifically propose the robust and accurate CF estimation (RACE) algorithm, which is mainly based on the mean-shift algorithm [20,21] and a two-direction-combination strategy. The mean-shift algorithm offers automatic, accurate and fast CF estimation, and the additional two-direction-combination strategy makes the CF estimation be further more accurate and robust. As a positive result, the motion estimation performance of SinMod would be promoted by optimizing it with RACE, and the correct assessment of cardiac function based on MRIT could be accordingly advanced.

2. Background

2.1. A general review of SinMod

The SinMod method proposed by Arts et al. [15] was designed for mapping displacement and deformation of the heart from 2-D TCMR images. As mentioned above, its main idea is to model the intensity at a pixel as the summation of two moving local sine wave-fronts. In this way, the local myocardial motion at each pixel is decomposed into two components, each one of which is along a wave-direction, and can be calculated from the frequency and phase-shift of the corresponding local sine-wave. The motion estimation is then converted to obtaining the local frequency and local phase-shift at each pixel.

Generally, to achieve the goal, SinMod works by extracting the relevant frequencies around a fundamental peak in the frequency domain, which is associated with a wave-direction, with two specifically designed band-pass filters, and generating the local frequency and local phase-shift maps with some calculations. As a consequence, the pixel-level MDF along that associated wave-direction can be derived. To estimate the overall motion, the same processing is repeated for another wave-direction, and the two resulting MDFs are added using the vector summation principle so as to form the final MDF, which consists of vectors that indicate the local motion displacements from one TCMR image to the other at corresponding pixel locations.

Specifically, for each wave-direction, the CF is first determined. According to Arts et al. [15], the CF refers to the spatial frequency of tags along the corresponding tag-direction, and is supposed to locate at a fundamental peak in the frequency spectrum of the TCMR image. For ease of expression, assuming that the CF has the location at frequency-domain coordinates $(\omega_{cx}, \omega_{cy})$, we denote it by a 2-D vector $\omega_c = (\omega_{cx}, \omega_{cy})$ (accordingly called the wave vector in Ref. [15]) with the amplitude $\omega_c = \|\omega_c\| = \sqrt{\omega_{cx}^2 + \omega_{cy}^2}$, where $\|\cdot\|$ denotes the norm operator. Using this CF, a basic band-pass filter b_{bf} is built by Eq. (1) and is tuned into a low-frequency filter b_{bflf} and a high-frequency filter b_{bfhf} by Eq. (2), where (ω_p, ω_q) denote the rotated coordinates that are converted from the original coordinates denoted by (ω_x, ω_y) , and i denotes the imaginary unit.

$$\begin{cases} b_{bf}(\omega_p, \omega_q) = \begin{cases} \cos^2\left(\frac{\pi Z}{2}\right), & Z < 1 \\ 0, & Z \geq 1 \end{cases} \\ Z = \left| \ln\left(\frac{\omega_p + i\omega_q}{\omega_c}\right) \right|. \end{cases} \quad (1)$$

$$\begin{cases} b_{bflf}(\omega_p, \omega_q) = \sqrt{\frac{\omega_c}{\omega_p}} b_{bf}(\omega_p, \omega_q), \\ b_{bfhf}(\omega_p, \omega_q) = \sqrt{\frac{\omega_p}{\omega_c}} b_{bf}(\omega_p, \omega_q), \\ (\omega_p, \omega_q) = \frac{1}{\omega_c} (\omega_x, \omega_y) \begin{pmatrix} \omega_{cx} & -\omega_{cy} \\ \omega_{cy} & \omega_{cx} \end{pmatrix}. \end{cases} \quad (2)$$

With the two filters b_{bflf} and b_{bfhf} , four filtered images I_{bflf1} , I_{bflf2} , I_{bfhf1} , I_{bfhf2} with complex intensities are derived from the two selected TCMR images. A low-frequency power image P_{lf} , a high-frequency power image P_{hf} and a cross-power image P_{cc} are then generated by Eq. (3). Finally, at a spatial location (p, q) , the local frequency $\omega_{\bar{p}}(p, q)$, the local phase-shift $\phi_{\bar{p}}(p, q)$ and the aimed local myocardial motion displacement $u_{\bar{p}}(p, q)$ along \bar{p} wave-direction are orderly calculated by Eq. (4), where \arg denotes the argument operator.

$$\begin{cases} P_{lf} = |I_{bflf1}|^2 + |I_{bflf2}|^2, \\ P_{hf} = |I_{bfhf1}|^2 + |I_{bfhf2}|^2, \\ P_{cc} = I_{bflf1} \bar{I}_{bflf2} + I_{bfhf1} \bar{I}_{bfhf2}. \end{cases} \quad (3)$$

$$\begin{cases} \omega_{\bar{p}}(p, q) = \omega_c \sqrt{\frac{P_{Hf}(p, q)}{P_{Lf}(p, q)}}, \\ \phi_{\bar{p}}(p, q) = \arg(P_{CC}(p, q)), \\ u_{\bar{p}}(p, q) = \frac{\phi_{\bar{p}}(p, q)}{\omega_{\bar{p}}(p, q)}. \end{cases} \quad (4)$$

The overall flowchart of the SinMod method is illustrated in Fig. 1, where the stage of CF estimation is specifically labeled. Due to its superior performance over HARP, SinMod has been studied and applied in several researches [16–19]. In Ref. [16], it is validated that automatic estimation of myocardial rotation and transmural shear can be performed well by tracking the TCMR image sequence using SinMod. In Refs. [17,18], displacements of the artificially added dense virtual tag intersections are estimated by SinMod and are then interpolated with the multilevel b-splines (MBS) algorithm [22]. Motion estimation in this way is continuous and smooth because of MBS, and is also fast and accurate due to SinMod. Besides, Ref. [18] also indicates that SinMod performs better than HARP in most cases

with respect to a distance metric. In Ref. [19], SinMod is extended to the three-dimensional (3-D) case. The 3-D SinMod method still works well on the artificial data added with Gaussian noise.

2.2. The CF estimation problem in SinMod

From Eqs. (1) and (2), it is obviously seen that the CF is always required for calculation, as in the form of either the amplitude ω_c or the coordinates $(\omega_{cx}, \omega_{cy})$. As a consequence, the motion estimation result of SinMod is heavily determined by the actual selection of CF. Nevertheless, we have already found that different CF selections would accordingly cause different errors in motion estimation, and sometimes the differences can hardly be neglected. Unfortunately, as previously mentioned, none of the original study of the SinMod method or the ones related to it clearly presented a detailed and automatic algorithm for CF estimation so far as we known. In order to reduce the motion estimation error of SinMod, an algorithm that can effectively and efficiently solve the CF estimation problem is much needed.

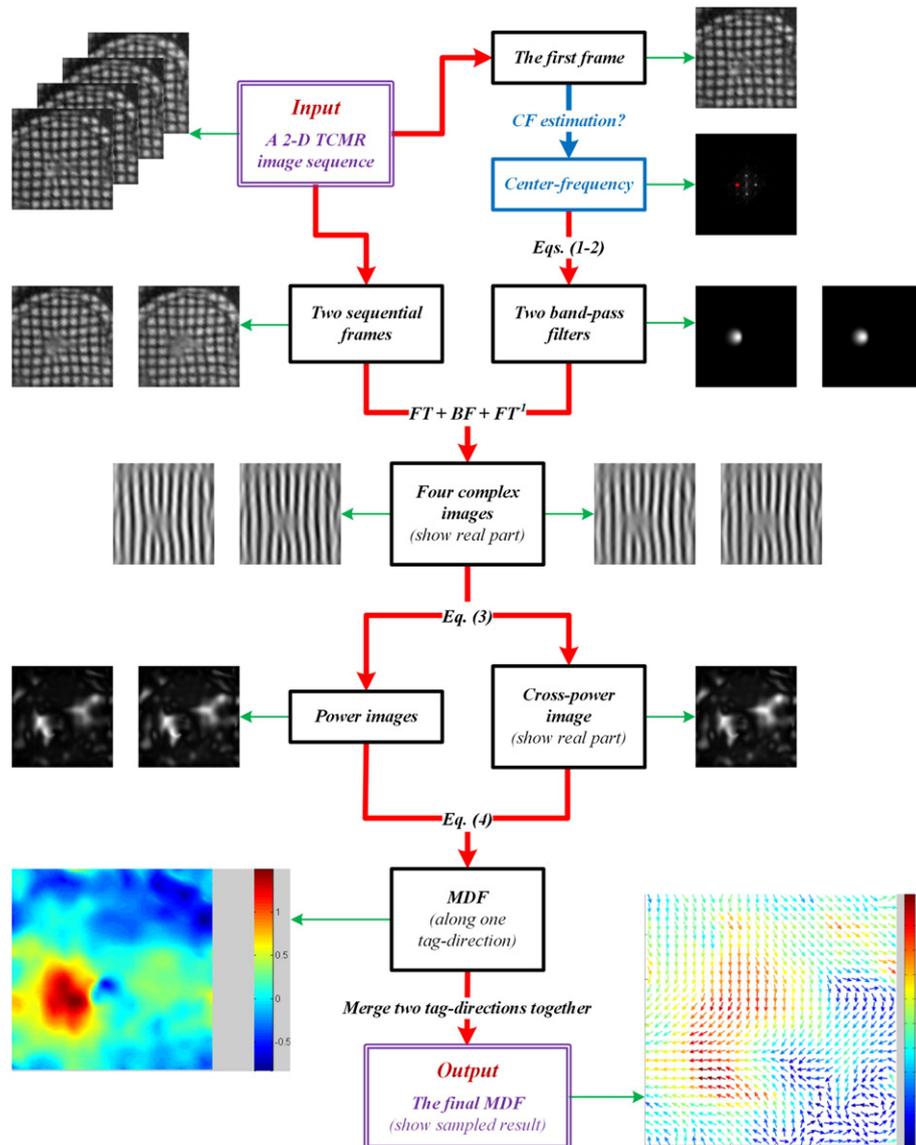


Fig. 1. The flowchart of the SinMod method. FT, BF and FT^{-1} denote Fourier transform, frequency-domain band-pass filtering and inverse Fourier transform, respectively.

The CF estimation problem can be formally described as: given a 2-D TCMR image sequence tagged in two tag-directions, a CF estimation algorithm is required, which can always provide an appropriate CF for the SinMod method, in order to make the cardiac motion estimation of that sequence using SinMod be as accurate as possible. According to the priori knowledge that whether the specified tagging parameters of the applied MRIT approach are known or not, the problem is dividedly analyzed in the following two cases.

2.2.1. With known tagging parameters

As illustrated in Fig. 2, in this case, the four tagging parameters including the initial tag spacings d_1 and d_2 , and the initial tag-direction angles θ_1 and θ_2 , are all given besides the TCMR sequence which is tagged in the two tag-directions t_1 and t_2 . According to the SinMod method, at a spatial location (p, q) , the same two wave-directions w_1 and w_2 are derived from t_1 and t_2 , respectively, using the perpendicularity relation. And for each w_i , ($i = 1, 2$), the ground truth of CF is accordingly located at either one of the two fundamental peaks f_{i1} and f_{i2} , which both correspond to w_i and are perfectly centrosymmetric about the frequency-domain origin o . As a result, there are totally four candidate CFs ω_{cij} , ($i, j = 1, 2$) that can be computed by Eq. (5), where row and col denote the row and column widths of each TCMR image, respectively.

$$\begin{cases} \omega_{c11} = (\omega_{cx11}, \omega_{cy11}) = \frac{1}{d_1} (\text{col} \sin\theta_1, -\text{row} \cos\theta_1), \\ \omega_{c12} = (\omega_{cx12}, \omega_{cy12}) = \frac{1}{d_1} (-\text{col} \sin\theta_1, \text{row} \cos\theta_1). \end{cases} \quad (5)$$

Due to the imaging protocols of the applied MRIT approach, such as the most common one, the spatial modulation of magnetization (SPAMM) [23,24], we should have $d_2 = d_1$ and $t_2 \perp t_1$ (i.e., $\theta_2 = \theta_1 \pm \pi/2$). Besides, row and col are also the same unless for special purpose. All these relations simplify Eq. (5) to Eq. (6), which indicates that the four candidate CFs are completely rotational symmetric about o with an interval angle of 90° .

$$\begin{cases} \omega_{c11} = \frac{\text{row}}{d_1} (\sin\theta_1, -\cos\theta_1), \\ \omega_{c12} = \frac{\text{row}}{d_1} (-\sin\theta_1, \cos\theta_1), \\ \omega_{c21} = \frac{\text{row}}{d_1} (\cos\theta_1, \sin\theta_1), \\ \omega_{c22} = -\frac{\text{row}}{d_1} (\cos\theta_1, \sin\theta_1). \end{cases} \quad (6)$$

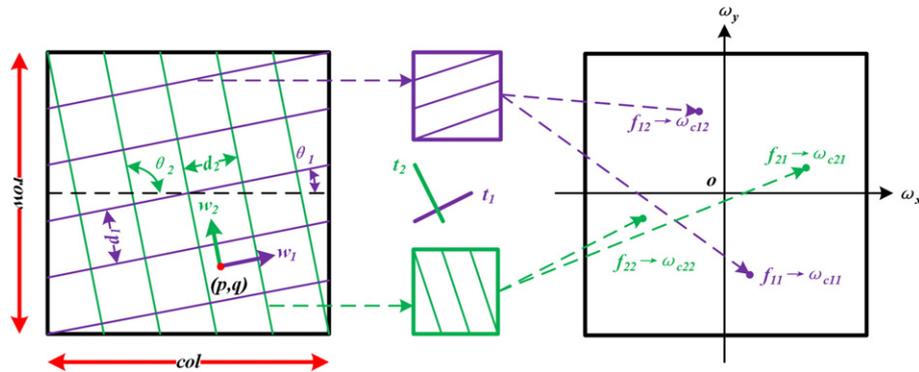


Fig. 2. The four tagging parameters $d_1, d_2, \theta_1, \theta_2$ in the spatial domain and the four candidate CFs ω_{cij} , ($i, j = 1, 2$) in the frequency domain. Left: a simulated image with tags along the two tag-directions t_1 and t_2 . Middle: the tags along each tag-direction. Right: the corresponding Fourier frequency spectrum. (ω_x, ω_y) and o denote the frequency-domain coordinates and origin (shifted to the spectrum center), respectively.

The above solution actually provides a way to compute the ground truth of CF. With this solution, as long as the parameters d_1, d_2, θ_1 and θ_2 are known, the CF can be estimated without error.

2.2.2. Without known tagging parameters

In this case, we only have the TCMR image sequence. A solution for the CF estimation problem seems theoretically feasible: estimating the four tagging parameters from the sequence beforehand to make Eqs. (5) and (6) be still available. However, the parameter estimation work is quite complicated and may not be worthy of doing with respect to its cost. In fact, we have tried some frequency estimation methods [25–27] which have good performance on one-dimensional signals, but they all failed in the parameter estimation probably because of the mutual interference between the two tag-directions. Hence, an effective way to solve the CF estimation problem without known tagging parameters is very necessary. Fortunately, Arts et al. [15] have implied three significant criterions that are all instructive and meaningful.

- (1) **Criterion 1: CF estimation for the first wave-direction.**
Here the “first” and “second” are used to distinguish the two wave-directions for the 2-D case: the first wave-direction is defined as perpendicular to the tag-direction, along which the tags have greater powers (measured in the frequency domain) than the tags along the other tag-direction. And so to define the second wave-direction. For the first wave-direction, Ref. [15] suggests that the CF (called the first CF to differentiate) is chosen “around the spatial frequency of the tags” along the corresponding tag-direction. As a result, the first CF is determined by “localization of regions of maximum power in the frequency domain”. Due to the strong periodicity of tags, this criterion is well founded, but note that it just gives a general thought rather than a specific solution.
- (2) **Criterion 2: CF estimation for the second wave-direction.**
For the second wave-direction, it is also introduced in Ref. [15] that the CF (called the second CF accordingly) is chosen “perpendicular to the first one with the same ω_c value.” In other words, the second CF is the rotated version of the first one around the frequency-domain origin with a rotation angle of 90° , where either clockwise or anticlockwise rotation is fine since both ways lead to the same result. Apparently, this criterion is theoretically logical according to Eq. (6). Although the presented tag spacings and tag-direction angles in the real TCMR image may be already changed due to the tag distortion and attenuation, the relation between the ground truths of two CFs would not be affected.

(3) **Criterion 3:** CF estimation for the whole sequence.

For the whole time-ordered TCMR sequence, it is not declared but could be speculated from Ref. [15] that, both CFs are estimated once from the first frame of the sequence, and are then shared with all the other frames. This criterion is also reasonable. Based on the principle of MRIT, among all the frames throughout a whole cardiac cycle, tags in the first frame should be distorted and attenuated the least. Consequently, each of the two CFs estimated from the first frame should be more approximate to the corresponding ground truth than the one estimated from any other frame. Moreover, as a matter of fact, CF estimation of the latter frames following Criteria 1–2 directly is very likely to fail, only if the distorted and attenuated tags are too weak to form obvious peaks in the frequency domain.

Acquiescently, a qualified CF estimation algorithm for the SinMod method should fully follow Criteria 1–3.

3. Methods

3.1. Some available CF estimation algorithms

Among Criteria 1–3, it is obvious that Criterion 1 is the determinant one for designing an eligible CF estimation algorithm, while Criteria 2–3 can be naturally followed. Hence, the key step of CF estimation is to effectively estimate the first CF from the first frame of the sequence. Actually, the following two algorithms can be logically put forward according to Criterion 1.

3.1.1. The MAX algorithm

The MAX algorithm may be the simplest algorithm that can fulfill the estimation of the first CF. In MAX, the frequency with the maximum power in the frequency domain, which accordingly locates at the brightest pixel in the Fourier power spectrum image, is taken as the first CF. Note that the direct current (DC) component of the power spectrum image should be excluded during the maximum seeking procedure in order to avoid mistake.

Apparently, the MAX algorithm has an extreme processing speed, but at the price that only integer coordinates can be obtained for the first CF. However, according to Eqs. (5) and (6), the ground truth of the first CF is supposed to have float coordinates. Therefore, the CF estimation accuracy of MAX is rough, which would cause an uncertain influence on the final motion estimation result.

3.1.2. The TRV algorithm

The TRV algorithm can be regarded as a major revision of MAX, and it is also more in line with Criterion 1. In TRV, a candidate region meaning to include all possible CF locations is first determined around the output of the MAX algorithm. After that, a moving fixed-size window is placed on the power spectrum image, each time with its center located at a different place within the candidate region. For every traveling position of that moving window, the powers of all frequencies fallen into it are added up using a weight function (can be chosen as Gaussian), and the result is assigned to the current window center. Finally, the center with the maximum power summation is taken as the location of the first CF.

The obvious advantage of TRV over MAX is that a CF estimate with float coordinates can be received, as long as the coordinates of chosen window centers within the candidate region are not limited to be integer, which can be easily carried out by setting the traveling step of the window as a small decimal less than one pixel. As a consequence, the CF estimation accuracy of TRV is absolutely higher than MAX.

However, TRV is much more complicated than MAX. Let the radius of the candidate region be r_{cnd} , the radius of the moving

window be r_{wnd} and the traveling step be $1/r_{trv}$, the time complexity of TRV will be $O(row \cdot col + r_{cnd}^2 r_{wnd}^2 r_{trv}^2)$. In contrast, that of MAX is only $O(row \cdot col)$. Even worse, since the precision of TRV which partly determines the CF estimation accuracy is completely controlled by r_{trv} , higher accuracy could be acquired by enlarging r_{trv} . Consequently, that would result in a much higher time complexity. Hence, there is a heavy contradiction between the high estimation accuracy and low time cost in TRV. Based on this reason, the TRV algorithm is also not recommended, although it seems very logical.

3.2. The RACE algorithm

We propose the RACE algorithm in this study aiming to solve the CF estimation problem in an automatic, accurate and fast way without known tagging parameters mentioned above. As previously introduced, the algorithm primarily utilizes the mean-shift algorithm and a two-direction-combination strategy. To be clear, we first analyze in theory that why both of the two techniques are applied for CF estimation, and then describe the detailed implementation of the RACE algorithm.

3.2.1. Theoretical analysis

(1) Reason for applying the mean-shift algorithm.

The mean-shift algorithm proposed by Fukunaga and Hostetler [20] is a well-known mode seeking algorithm that has already been applied in many fields, including computer vision and image processing [21]. Given a set of discrete data sampled from a density function and an appropriate starting point, mean-shift can automatically locate the maximum density of that function with both high accuracy and fast speed by means of an iteration procedure.

With respect to the CF estimation problem, we find that the mean-shift algorithm is very suitable. First, the power spectrum image can absolutely be taken as a 2-D discrete density function, which shows the power magnitude of each frequency as the number of coincided data points. Second, according to Criterion 1, the maximum density of the local region around a fundamental frequency peak is highly accordant with the ground truth of the corresponding CF. Furthermore, the high accuracy and fast speed which mean-shift can concurrently guarantee are both in great demand.

(2) Reason for applying the two-direction-combination strategy.

According to Criterion 3, for the whole sequence, the two CFs are estimated from the first frame. Since in the first frame, the tags are unavoidably distorted and attenuated to a minor extent; on the other hand, the heterogeneity of underlying tissue intensities and various noises also disturb the tag periodicity. As a consequence, it is likely that the estimate of the first CF would be already deviated from the ground truth. With a direct application of Criterion 2, the deviation would still remain between the estimate of the second CF and its corresponding ground truth, as illustrated in Fig. 3A.

The essential cause of the above problem is that, Criterion 2 only emphasizes the constraint relation between the two CFs, but neglects the independent acquisition feasibility of the second CF. In fact, the second CF can also be estimated in the same way as the first one. In this fashion, the deviation between each CF estimate and its ground truth is generally different from the other deviation (see Fig. 3A). Or further, the two deviations may be able to be mutually neutralized to a certain extent, since distortion and attenuation of tags along one tag-direction is partially independent from the other.

Following the above inspiration, we technically develop the two-direction-combination strategy, the main idea of which is to combine the two wave-directions together for a conjoined estimation. With

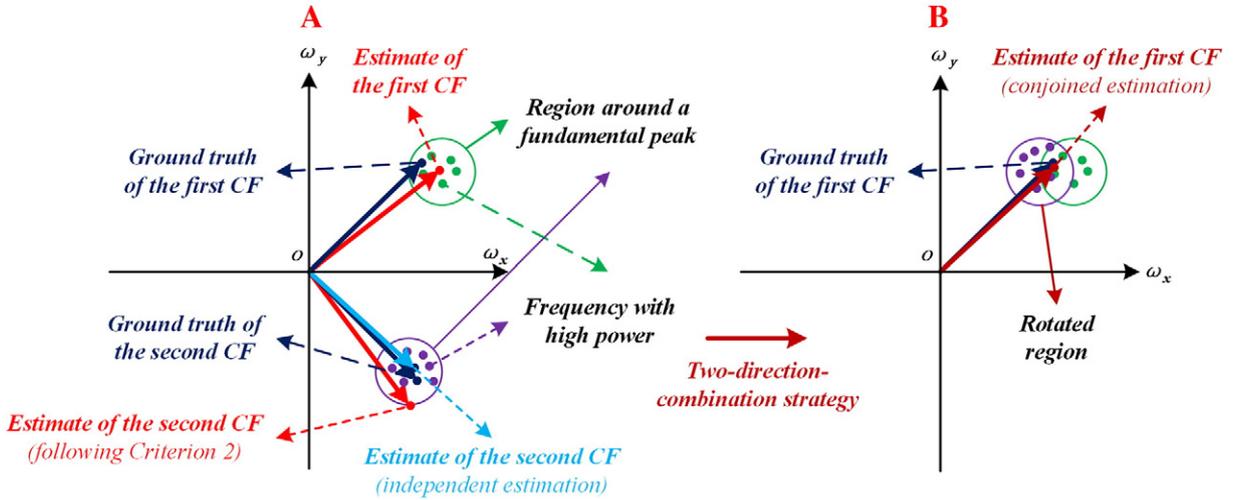


Fig. 3. The role of the two-direction-combination strategy. A. CF estimation directly following Criterion 2/independent CF estimation. B. Conjoined CF estimation using the strategy.

this strategy, two benefits could be gained. First of all, it is guaranteed that the estimation error of either CF will at least not exceed the larger error of the two CFs that are independently estimated. Theoretically, there is a 50% probability that the conjoined estimation would achieve the least error. Second, due to the underlying tissue intensities and noises, the powers of frequencies within the mean-shift region are affected in different degrees. According to high similarity of tags and poor similarity of other influential factors, the conjoined approach is supposed to enhance the powers of those frequencies close to the ground truth of CF, and meanwhile suppress the affection caused by most irrelevant factors. In conclusion, it is very logical and convincible that this strategy can effectively improve the accuracy (see Fig. 3B) and robustness of CF estimation.

3.2.2. Implementation

In the current implementation, the proposed RACE algorithm primarily includes the following five key steps.

(1) Normalize all frames of the sequence.

First of all, every frame of the TCMR sequence is normalized in the same way: mean to 0 and root mean square (RMS) to 1, as well as in Ref. [15]. With this preprocessing, some global intensity variations happening to the sequence can be generally corrected. More importantly, the normalization can totally eliminate the DC component of the Fourier power spectrum image, which greatly helps in correctly locating the frequency with the maximum power as the MAX algorithm does.

(2) Generate the revised power spectrum image.

Specifically, the two-direction-combination strategy is carried out by replacing the Fourier power spectrum image with a revised one in the first place. To effectively fulfill the strategy, we define the revised power spectrum image as the summation of the original one and its rotation around the frequency-domain origin with an angle of 90° . In this way, the resulting image is rotational symmetric around its center, and each wave-direction can be taken for the estimation of the first CF.

(3) Initialize the mean-shift region.

We set up the mean-shift region with a circular envelope for perfect symmetry, a shifting shape-center determined by an initial position and the mean-shift iteration procedure, and a varying radius which is kept to be half of the distance between

the shape-center and the frequency-domain origin for excluding irrelevant information. The only one parameter to be required is the initial shape-center. Fortunately, we can directly set it to locate at the frequency with the maximum power, just like the way MAX performs. Thus, the RACE algorithm can hold its automaticity and will not lose the correctness.

(4) Compute the mass-center.

Following the principle of the basic mean-shift algorithm, we compute the mass-center $\omega_{mc} = (\omega_{mcx}, \omega_{mcy})$ by Eq. (7), where R_{ms} denotes the mean-shift region, $\omega_{sc} = (\omega_{scx}, \omega_{scy})$ denotes the shape-center of R_{ms} , r_{ms} denotes the radius of R_{ms} , ω denotes a frequency in R_{ms} , $P_{revised}$ denotes the revised power spectrum image, and $K_{2D-Gaussian}$ denotes the kernel function currently chosen as the 2-D Gaussian kernel [28]. By keeping replacing ω_{sc} with ω_{mc} , Eq. (7) is repeatedly performed until the two centers completely coincide.

$$\left\{ \begin{array}{l} \omega_{mc} = \frac{\sum_{\omega \in R_{ms}} K_{2D-Gaussian}(\omega - \omega_{sc}) P_{revised}(\omega) \omega}{\sum_{\omega \in R_{ms}} K_{2D-Gaussian}(\omega - \omega_{sc}) P_{revised}(\omega)} \\ R_{ms} = \{\omega \mid \|\omega - \omega_{sc}\| \leq r_{ms}\}, r_{ms} = \frac{\|\omega_{sc}\|}{2}, \\ K_{2D-Gaussian}(\omega) = \frac{1}{2\pi\sigma^2} e^{-\frac{\|\omega\|^2}{2\sigma^2}}, \sigma = \frac{r_{ms}}{3}. \end{array} \right. \quad (7)$$

(5) Produce the two CF estimates.

Once the iteration of Eq. (7) terminates, the final ω_{mc} is taken as the first CF estimate $\omega_{est1} = (\omega_{est1x}, \omega_{est1y})$, and the second CF estimate ω_{est2} is computed from ω_{est1} using Eq. (8), as directly following Criterion 2.

$$\omega_{est2} = \omega_{est1} \begin{pmatrix} 0 & \pm 1 \\ \mp 1 & 0 \end{pmatrix} = (\mp \omega_{est1y}, \pm \omega_{est1x}) \quad (8)$$

The flowchart of the RACE algorithm is illustrated in Fig. 4, where the two stages that utilize the mean-shift algorithm and the two-direction-combination strategy, respectively, are specifically labeled. As can be proved, the time complexity of RACE is less than $O(\text{row} \cdot \text{col} + n_{mit}(\max(r_{ms}))^2)$, where n_{mit} denotes the maximum iteration. According to abundant experiments, n_{mit} seldom surpasses 5 times. Obviously, FACE runs much faster than TRV.

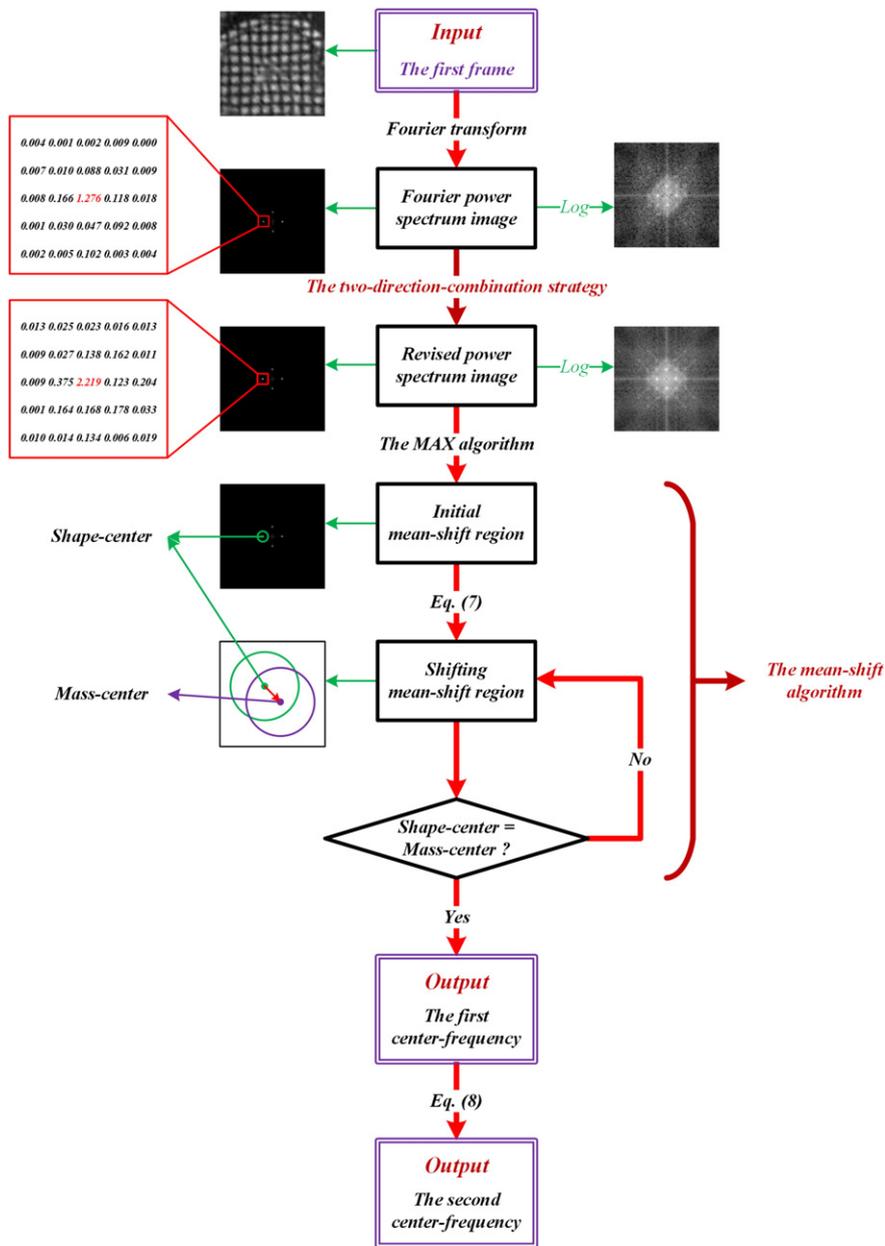


Fig. 4. The flowchart of the RACE algorithm. Red rectangles both filled with 5×5 data show the local power distributions of corresponding power spectra with the same magnitude of 10^7 .

4. Experiments

4.1. Materials

Our TCMR data is collected from a cooperative hospital. All data was acquired on a 1.5 T GE MR scanner (Signa HDxt, GE Healthcare, Milwaukee, Wisconsin, USA). In this study, two groups of human in vivo TCMR images with high contrast are specifically chosen for experiments. The first image group consists of 3 sequences (numbered from 1 to 3) corresponding to three sequential cardiac slices from a healthy volunteer, while the second image group consists of 4 sequences (numbered from 4 to 7) corresponding to four sequential cardiac slices from a patient with cardiomyopathy.

For the healthy volunteer, the primary imaging parameters were set as: repetition time (TR) = 6.7 ms, echo time (TE) = 3.2 ms, field of view (FOV) = 32 cm \times 32 cm, flip angle (FA) = 12°, acquisition matrix = 256 \times 160, slice thickness = 8 mm, spacing between slices = 10 mm,

cardiac number of images = 20, rows = 256, columns = 256; for the patient, the parameters were accordingly set as: TR = 6.6 ms, TE = 3.1 ms, FOV = 36 cm \times 36 cm, FA = 12°, acquisition matrix = 256 \times 160, slice thickness = 8 mm, spacing between slices = 11 mm, cardiac number of images = 20, rows = 256, columns = 256.

Consequently, each sequence includes 20 frames, which all have the same size of 256 \times 256 pixels and together represent a whole cardiac cycle. To facilitate studying, each frame is rotated 45° anticlockwise to make the tags be presented in both horizontal and vertical directions. To highlight the left ventricle which is taken as the region of interest, all frames are cut down to a new size of 110 \times 110 pixels with the same cutting position. For the purpose of visualization, the 1st, 6th, 11th, 16th and 20th frames of each of the 7 sequences are shown in Fig. 5.

For the running environment, our experiments are all performed in the R2014a 64bit version of MATLAB software with the hardware of 4 \times 2.7 GHZ processors and 8 GB RAM.

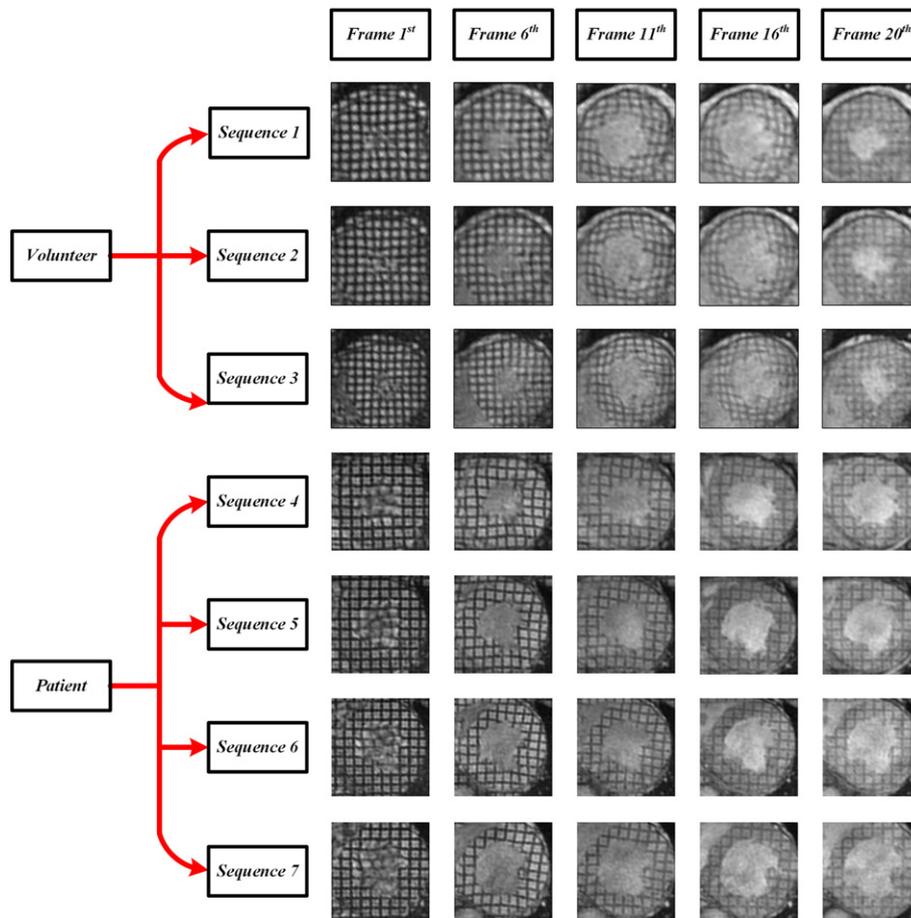


Fig. 5. The TCMR data (partly shown) chosen for experiments.

4.2. Comparison methods

In our experiments, the TRV algorithm is abandoned due to its extraordinary high time cost, and the following three CF estimation algorithms are specifically chosen for comparison.

- (1) The RACE algorithm.
- (2) The MAX algorithm.
- (3) The MS algorithm, which utilizes the mean-shift algorithm yet without the two-direction-combination strategy. Note that the only difference between MS and FACE is the selection of the power spectrum image.

The performances of above three algorithms are compared in the aspect of CF estimation result. In addition, each algorithm is separately integrated into the original SinMod method, yielding three revised methods that are denoted by SinMod & RACE, SinMod & MAX and SinMod & MS, respectively. And the performances of these methods are compared in the aspect of MDF estimation result.

4.3. Validation approaches

To reasonably validate the CF/MDF estimation result of each algorithm/method, it would be very helpful if the corresponding ground truth is already known. However, for real in vivo cardiac data, that requirement can hardly be satisfied. To address this “validation without ground truths” problem, several categories of effective solutions have been investigated in past decades, including (1) manufacturing simulated data with the known ground truths [29–31], (2) artificially marking out ground truths [32], (3) using

gold-standard methods to generate ground truths [29,31], and (4) other particular approaches [33]. Nevertheless, these solutions all have limitations.

As replacements, in this study, we specifically bring out the following validation approaches, which we consider to be more impersonal and impartial for evaluation and comparison.

4.3.1. Validation for CF estimation

According to Eqs. (5) and (6), for each CF, as long as the corresponding tag spacing and tag-direction angle are both unaltered, its ground truth will remain the same. This fact provides a creative way for the validation of CF estimation bypassing the necessity of the ground truth: cautiously altering the TCMR image without changing the tag spacings and tag-direction angles, and weighing the variance of the multi CF estimates. Ideally, the variance should perfectly be zero. Regarding the actual situation, although the variance can hardly remain zero, it is still factual that the smaller the variance, the more robust the CF estimation would be.

To ensure that the tag spacings and tag-direction angles do not change notwithstanding the image altering, the following two approaches are designed.

- (1) Adding different levels of noise to the image subject.
- (2) Assigning the image subject to be the cutting result of a larger-sized source image. In this way, the image subject varies with different cutting positions, which however only causes translation motion to the tags.

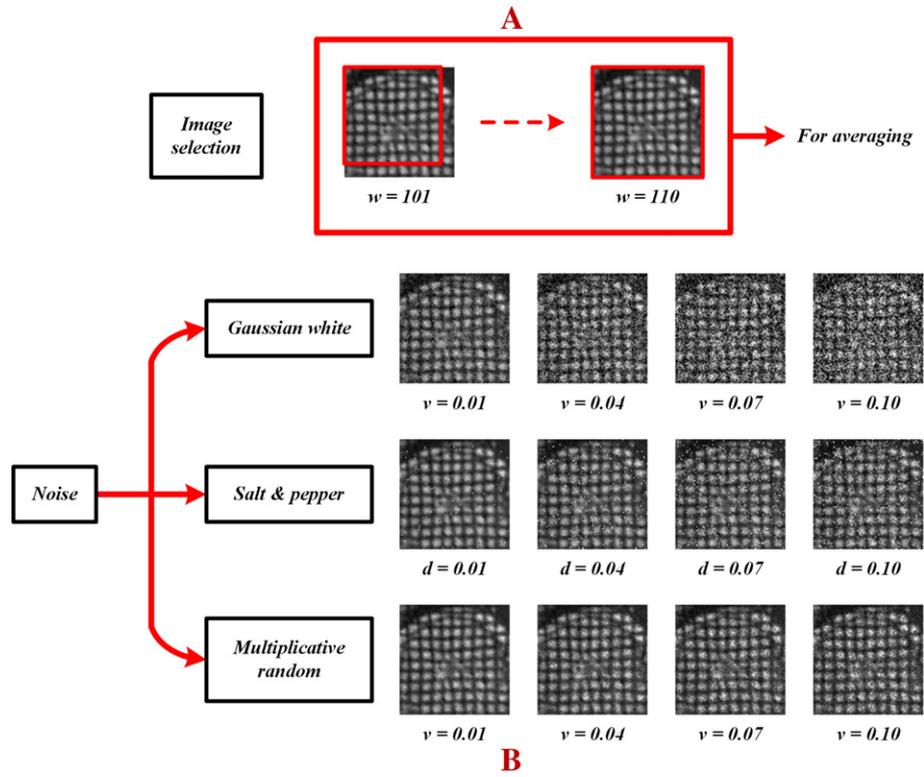


Fig. 6. A. The image selection scheme of the first experiment. B. The TMCR images noised with different noise types and levels. w , v and d denote the image width, noise variance and noise density, respectively.

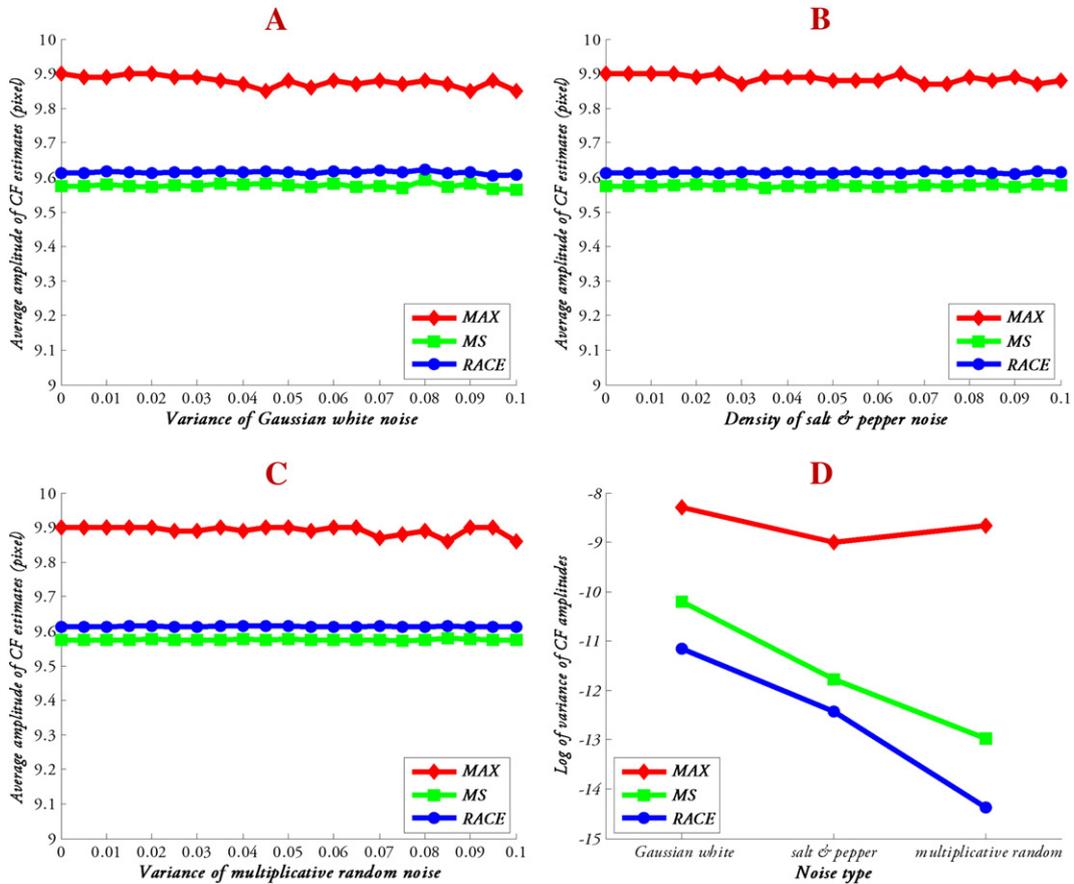


Fig. 7. Results of the first experiment. A–C. Average amplitude of CF estimates vs. noise level for the three types of noises, respectively. D. Log of CF amplitudes.

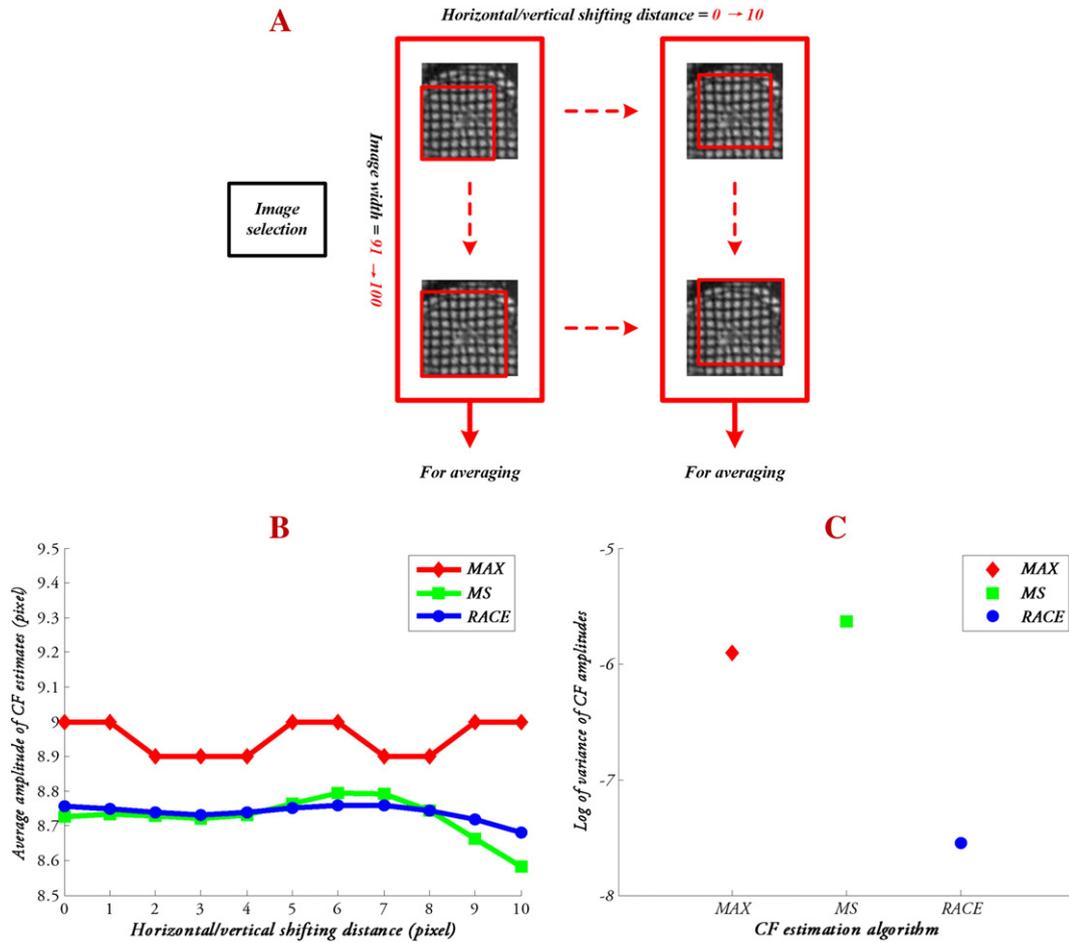


Fig. 8. A. The image selection scheme of the second experiment. B. Average amplitude of CF estimates vs. shifting distance. C. Logs of CF amplitudes.

Note that to eliminate non-uniqueness, we choose the CF amplitude for variance measurement, since both of the first and the second CFs have just the same amplitude.

4.3.2. Validation for MDF estimation

As a matter of fact, the ground truth of MDF between two real TCMR images is also impossible to be directly acquired, but it is entirely feasible to create a known ground truth of MDF on condition that the two image subjects are generated from the same source image. Based on this consideration, two approaches are specifically designed for the validation of MDF estimation.

- (1) Assigning the second image subject to be a shifted version of the first one. In this way, the ground truth of MDF between them is known as the global translation.
- (2) The two image subjects are not directly associated, but once the MDF between them is estimated, the second image is replaced by one of its shifted versions, and the same motion estimation is performed again. As a result, the ground truth of the difference between the two MDFs is known as the global translation as well.

For accuracy measurement, the mean error and the RMS error are both calculated. Concretely, in the first approach, let the ground truth of MDF be U_{tru} and the estimated MDF be U_{est} , the two errors denoted by ϵ_{mean} and ϵ_{rms} , respectively, are computed by Eq. (9), where R_{stat} denotes the statistics region used for excluding the less concerned

image parts that without distinct tags, and (i, j) denotes an image pixel within R_{stat} . In the second approach, let the first and the second estimated MDFs be U_{est1} and U_{est2} , respectively, and the ground truth of the difference between two MDFs be $U_{\Delta tru}$, the two errors denoted by $\epsilon_{\Delta mean}$ and $\epsilon_{\Delta rms}$, respectively, are computed by Eq. (10).

$$\left\{ \begin{aligned} \epsilon_{mean} &= \frac{\sum_{(i,j) \in R_{stat}} \|U_{est}(i, j) - U_{tru}(i, j)\|}{Area(R_{stat})}, \\ \epsilon_{rms} &= \sqrt{\frac{\sum_{(i,j) \in R_{stat}} \|U_{est}(i, j) - U_{tru}(i, j)\|^2}{Area(R_{stat})}}. \end{aligned} \right. \quad (9)$$

$$\left\{ \begin{aligned} \epsilon_{\Delta mean} &= \frac{\sum_{(i,j) \in R_{stat}} \|U_{est2}(i, j) - U_{est1}(i, j) - U_{\Delta tru}(i, j)\|}{Area(R_{stat})}, \\ \epsilon_{\Delta rms} &= \sqrt{\frac{\sum_{(i,j) \in R_{stat}} \|U_{est2}(i, j) - U_{est1}(i, j) - U_{\Delta tru}(i, j)\|^2}{Area(R_{stat})}}. \end{aligned} \right. \quad (10)$$

Note that we only create the global translation motion in this study. The reason we do not test more kinds of motion, such as the

Table 1

The processing time of each algorithm/method.

| | MAX (ms) | MS (ms) | RACE (ms) | TRV (s) | | | | | SinMod (ms) |
|------------|----------|---------|-----------|------------|------------|-------------|-------------|-------------|-------------|
| | | | | Step = 1/2 | Step = 1/5 | Step = 1/10 | Step = 1/20 | Step = 1/50 | |
| Sequence 1 | 0.73 | 0.93 | 1.40 | 0.051 | 0.273 | 1.081 | 4.356 | 27.219 | 7.15 |
| Sequence 2 | 0.66 | 0.97 | 1.51 | 0.062 | 0.392 | 1.532 | 6.179 | 38.255 | 7.30 |
| Sequence 3 | 0.67 | 1.01 | 1.40 | 0.113 | 0.533 | 2.122 | 8.485 | 53.092 | 7.41 |
| Sequence 4 | 0.68 | 0.97 | 1.43 | 0.030 | 0.235 | 0.739 | 2.970 | 18.566 | 6.74 |
| Sequence 5 | 0.68 | 0.91 | 1.31 | 0.045 | 0.323 | 1.085 | 4.341 | 27.064 | 6.68 |
| Sequence 6 | 0.65 | 1.00 | 1.41 | 0.061 | 0.382 | 1.523 | 6.137 | 38.159 | 6.95 |
| Sequence 7 | 0.66 | 0.87 | 1.27 | 0.075 | 0.307 | 1.078 | 4.316 | 27.153 | 6.52 |

global rotation and scaling, is to avoid large local motion and the possible errors caused by essential interpolation.

4.4. Experimental results

4.4.1. CF estimation

In the first experiment, we study the performance of each CF estimation algorithm by following the first validation approach of CF estimation. As illustrated in Fig. 6A, the first frame of Sequence 1 is chosen to generate the TCMR image subject using a square window, the width of which varies from 101 to 110 pixels in order to eliminate the dependence of the results to a specific ground truth of CF. Three common types of noises, including the Gaussian white noise, the salt & pepper noise and the multiplicative random noise, each with varying levels, are added into the image subject separately. Specifically, for the Gaussian white noise and multiplicative random noise, their variances both vary from 0.01 to 0.1 with a step of 0.005; for the salt & pepper noise, its density varies in the same way. Fig. 6B partly shows the noised images. And for each noise level and each image width, the image subject is repeatedly noised 10 times in order to suppress the randomness of noise, which totally yields 100 CF estimates in consideration of the 10 specified image widths, and the amplitudes of all these estimates are averaged to generate a statistical result. The relation curves of the average amplitude of CF estimates vs. the noise level (variance or density) of the three algorithms in regard to the three different types of noises are shown in Fig. 7A–C, respectively. And the corresponding logs of variance of CF amplitudes are calculated and also shown in Fig. 7D. It can be seen that the mean value of the average CF amplitudes always differs from one algorithm to another no matter for which hired noise. Since the corresponding ground truth is unknown, we cannot tell which algorithm estimates the most accurately. Yet the curves clearly show that for each noise type, the RACE algorithm always has the smallest variance of CF amplitudes, while the MAX algorithm continuously gets the largest one, and the MS algorithm keeps being close to but never beyond RACE (see Fig. 7D). According to all the results, we can generally draw the conclusion that in contrast with the other two algorithms, RACE has the strongest robustness of CF estimation with respect to noise.

In the second experiment, accordingly, we study the performance of each algorithm by following the second validation approach of CF

estimation. As illustrated in Fig. 8A, the first frame of Sequence 1 is still chosen to generate the image subject, only this time using a fixed-size square window, the location of which shifts diagonally from the bottom left to the top right of the source image. For the same purpose, the image width remains no change during a whole shifting round, but varies from 91 to 100 pixels to create 10 independent shifting rounds for averaging. Likewise, the relation curves of the average amplitude CF estimates vs. the horizontal/vertical shifting distance of the three algorithms are shown in Fig. 8B, and the corresponding logs of variance of CF amplitudes are shown in Fig. 8C. Once again, the RACE algorithm absolutely gains the smallest variance of CF estimates. And it is seen that the MAX algorithm has a smaller variance than MS this time, but note that the fluctuation of CF amplitudes estimated by MS is presented in a much smoother way than MAX, which also symbolizes a stronger robustness in some way. In general, the similar conclusion can be drawn: the RACE algorithm still has the strongest robustness of CF estimation among the three algorithms with respect to image translation.

In addition, Table 1 shows the processing times of all involved algorithms/methods, including MAX, MS, RACE, TRV (with different settings of the traveling step) and SinMod (with no CF estimation), performing on the first frame of each sequence using generally optimized MATLAB codes. From the table, it is seen that the three comparison algorithms MAX, MS and RACE all have very low time costs as compared with the main body of SinMod (only about 1/10 ~ 1/5 times), and they behave quite close to each other. Although the MAX algorithm always performs the fastest while RACE the slowest, the differences among them can be almost ignored (<1 ms). On the contrary, the TRV algorithm shows a much huger time cost. As the traveling step narrows down from 1/2 to 1/50 pixel, the cost grows exponentially from about 10 to about 5000 times of that of SinMod, which is definitely undesirable. Therefore, only from the perspective of time, the TRV algorithm deserves to be waived, while the other algorithms are quite optional.

And Table 2 accordingly shows the estimation results of the first CF of all algorithms on the first frame of each sequence. As can be seen, the MAX algorithm receives integer coordinates every time, while the others can all get float ones. Both of the MS and RACE algorithms are able to achieve very high precision, and their results are very close but never fully equal to each other. As for the TRV algorithm, with a large traveling step, its performance degrades to

Table 2

The estimation result of the first CF of each algorithm.

| | MAX | MS | RACE | TRV | | | | |
|------------|---------|--------------------|--------------------|------------|--------------|--------------|----------------|----------------|
| | | | | Step = 0.5 | Step = 0.2 | Step = 0.1 | Step = 0.05 | Step = 0.02 |
| Sequence 1 | (10, 0) | (9.9743, 0.0640) | (9.9403, 0.0913) | (10, 0) | (10.0, 0.2) | (10.0, 0.1) | (10.00, 0.05) | (10.00, 0.06) |
| Sequence 2 | (11, 0) | (10.7812, 0.0489) | (10.6441, 0.0392) | (11, 0) | (10.7, 0.1) | (10.8, 0.0) | (10.80, 0.05) | (10.78, 0.04) |
| Sequence 3 | (12, 0) | (12.0391, 0.0008) | (11.9762, -0.0357) | (12, 0) | (12.0, 0.0) | (12.1, 0.0) | (12.05, 0.00) | (12.04, 0.00) |
| Sequence 4 | (9, 0) | (9.0998, -0.0539) | (9.0743, -0.0167) | (9, 0) | (9.1, 0.1) | (9.1, -0.1) | (9.10, -0.10) | (9.10, -0.08) |
| Sequence 5 | (10, 0) | (9.9349, -0.0191) | (9.9126, -0.0429) | (10, 0) | (9.8, 0.0) | (9.9, 0.0) | (9.95, 0.00) | (9.92, -0.06) |
| Sequence 6 | (11, 0) | (10.9060, -0.0191) | (10.7532, -0.0457) | (11, 0) | (10.9, -0.1) | (10.9, -0.1) | (10.85, -0.00) | (10.84, -0.06) |
| Sequence 7 | (10, 0) | (9.9700, -0.0152) | (9.9511, -0.0509) | (10, 0) | (10, 0.0) | (9.9, 0.0) | (9.95, 0.00) | (9.98, 0.00) |

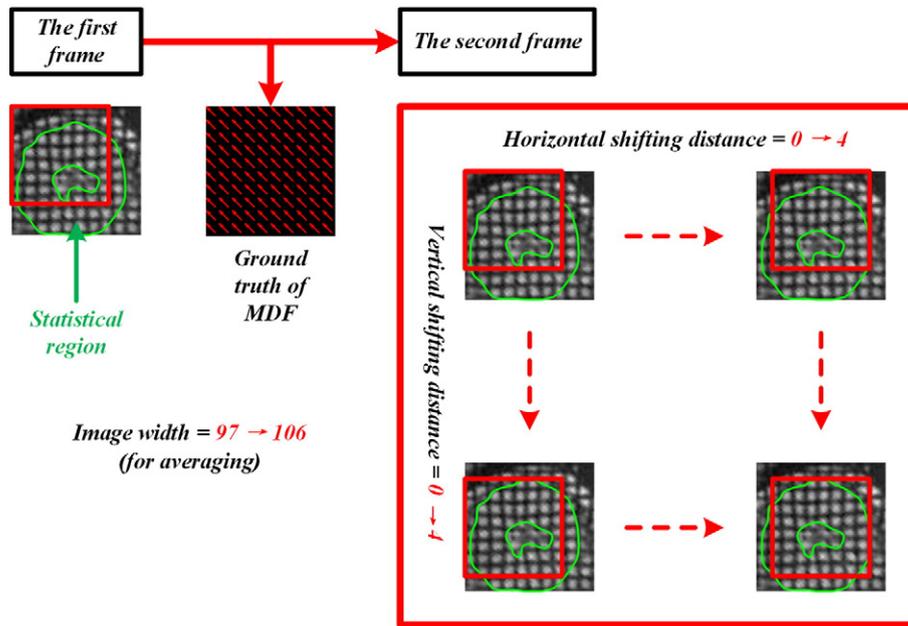


Fig. 9. The image selection scheme of the third experiment.

linearly and the estimation result tends to converge in an unstable way. Overall, all the four algorithms can yield similar results for each sequence. And from these results, it seems intuitively likely that the MS and RACE algorithms always get closer to the corresponding ground truths than the other two.

4.4.2. MDF estimation

In the third experiment, we study the performance of each revised method by following the first validation approach of MDF estimation. As illustrated in Fig. 9, the first TCMR image subject is fixed at the top left of the chosen source image, while the second image subject is sliding from the top left to the bottom right of the same source image with the maximum horizontal/vertical shifting distance of 4 pixels. And the green double contours outline the statistical region. Similar as the second experiment, the image width varies from 97 to 106 pixels outside a one-time shifting round for averaging. The relation curves of the average mean/RMS error of MDF estimation vs. the horizontal and vertical shifting distances of the three methods on the first frame of Sequence 1 are shown in Fig. 10A. It is clearly seen that in each shifting case, the SinMod & RACE method constantly achieves the smallest errors in both aspects, while the SinMod & MS method has a similar performance and SinMod & MAX behaves distinctly worse all the time. And as the shifting distance in either direction becomes larger, both errors tend to grow in an approximate linear and smooth way for each method. The results adequately stand for the superiority of the RACE algorithm in upgrading the motion estimation accuracy of the SinMod method over the other two algorithms.

Fig. 10B–C additionally shows the motion error fields and part of their differences in two specific cases for a clearer comparison. As shown in Fig. 10B, for the case with a smaller shifting distance of 2 pixels in the horizontal direction only, the difference of local motion errors between the SinMod & RACE method and each of the other two is not very obvious, but the superiority of the former method can be confirmed statistically. As shown in Fig. 10C, for another case with a larger shifting distance of 4 pixels in the vertical direction only, the corresponding differences become quite distinct in some local regions (see the two outlined ellipses in Fig. 10C), which visually indicate the ascendancy of the RACE algorithm.

Considering the possible individual difference coming from one single cardiac slice, Fig. 11A–B further shows the average curves of the three methods on Sequences 1–3 as corresponding to the volunteer and Sequences 4–7 as corresponding to the patient. Similarly, in almost every shifting case, the SinMod & RACE method wins the smallest errors. Specifically, for the mean error, SinMod & RACE always performs the best, SinMod & MS takes the second place with small differences and SinMod & MAX shows the worst performance with much larger gaps especially for the volunteer. For the RMS error, differently, all the three methods behave highly consistent with each other. In spite of that, it is still discernible that the SinMod & RACE method remains achieving the least errors with very few exceptions. Taken all together, it can be strongly concluded that as compared with MS and MAX, RACE can promote the motion estimation accuracy of the SinMod method to the highest level.

In the fourth experiment, we study the performance of each revised method by following the second validation approach of MDF estimation. As illustrated in Fig. 12, for the first MDF estimation, the first and the second image subjects are cut from the first and the second frames of the same chosen sequence, respectively, both at the top right positions; for the second MDF estimation, the first image stays the same, while the second image is sliding from the top right to the bottom left of the same source image. The maximum shifting distance and the variation range of image width are both the same as in the third experiment. Similarly, the relation curves of the average mean/RMS of MDF estimation difference vs. the shifting distances in both directions on Sequences 1–3 and Sequences 4–7 (in the form of average as well) are shown in Fig. 13A–B, respectively. The results present a high consistency with those in Fig. 11. Likewise, it can be also concluded that the RACE algorithm can promote the motion estimation robustness of the SinMod method to the strongest level as compared with the other two algorithms.

Table 3 further shows the relative mean/RMS error of each revised method on each sequence using the results of the third and fourth experiments. Here the relative error e is computed by Eq. (11), where d_{hrz} and d_{vrt} denote the horizontal and vertical shifting distances, respectively, and \bar{e} denotes the average error of all cases yielded from the image width variation. It is clearly indicated in Table 3 that the SinMod & RACE method always obtains the least errors expect for very few cases, although its advantage over the

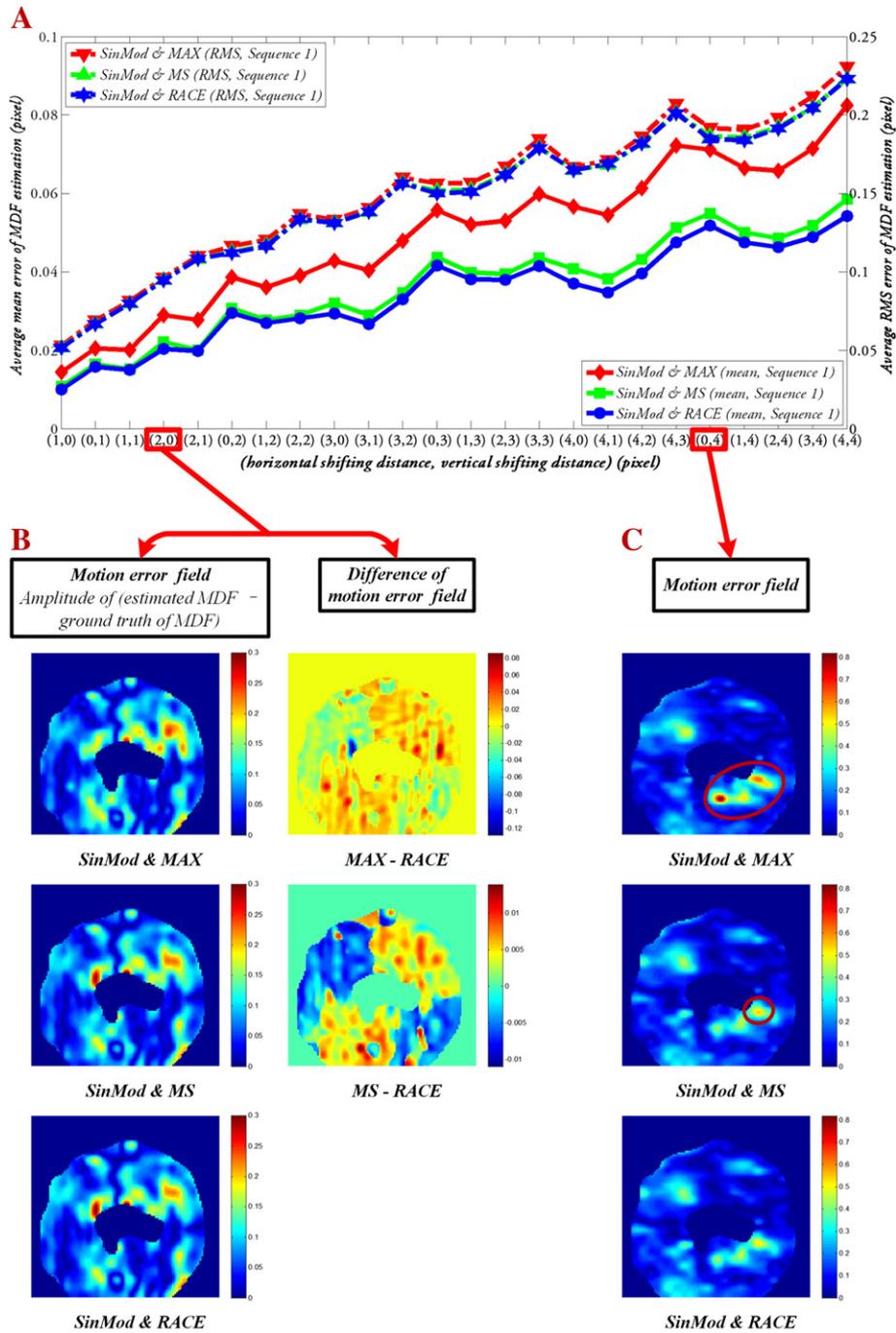


Fig. 10. A. Average mean/RMS error of MDF estimation vs. shifting distances for Sequence 1. B. Motion error fields and their differences in the shifting case of (2, 0). C. Motion error fields in the shifting case of (0, 4).

other two may not be very large. The averaged data also listed in the table gives a more obvious comparison.

$$\begin{cases} e_{ind} = \frac{1}{24} \sum_{\substack{0 \leq d_{hrz} \leq 4, 0 \leq d_{vrt} \leq 4 \\ d_{hrz} + d_{vrt} \neq 0}} \frac{\bar{e}_{ind}(d_{hrz}, d_{vrt})}{\sqrt{d_{hrz}^2 + d_{vrt}^2}}, \\ ind = mean, rms, \Delta mean, \Delta rms. \end{cases} \quad (11)$$

Finally, Fig. 14A–E shows the continuous frame-to-frame motion tracking results of the tags in the first five frames of Sequence 1,

respectively, using the SinMod & RACE method. The display area of each figure is set as the statistic region of the first frame. The green lines in each figure denote the tags directly detected from the underlying frame using the HARP method, while the red lines denote the tags tracked by SinMod & RACE and deformed by MBS. As shown in the figures, the two sets of tags are completely overlapping at the very beginning, and start to split as the frame number increases, however, in a very slow and local way. Actually, even in Fig. 14E, the overlap ratio of them is still high, and the local divergences mainly occur at the places close to the inner and outer contours of the statistic region. Therefore, the correctness of the RACE algorithm is proved to be well enough.

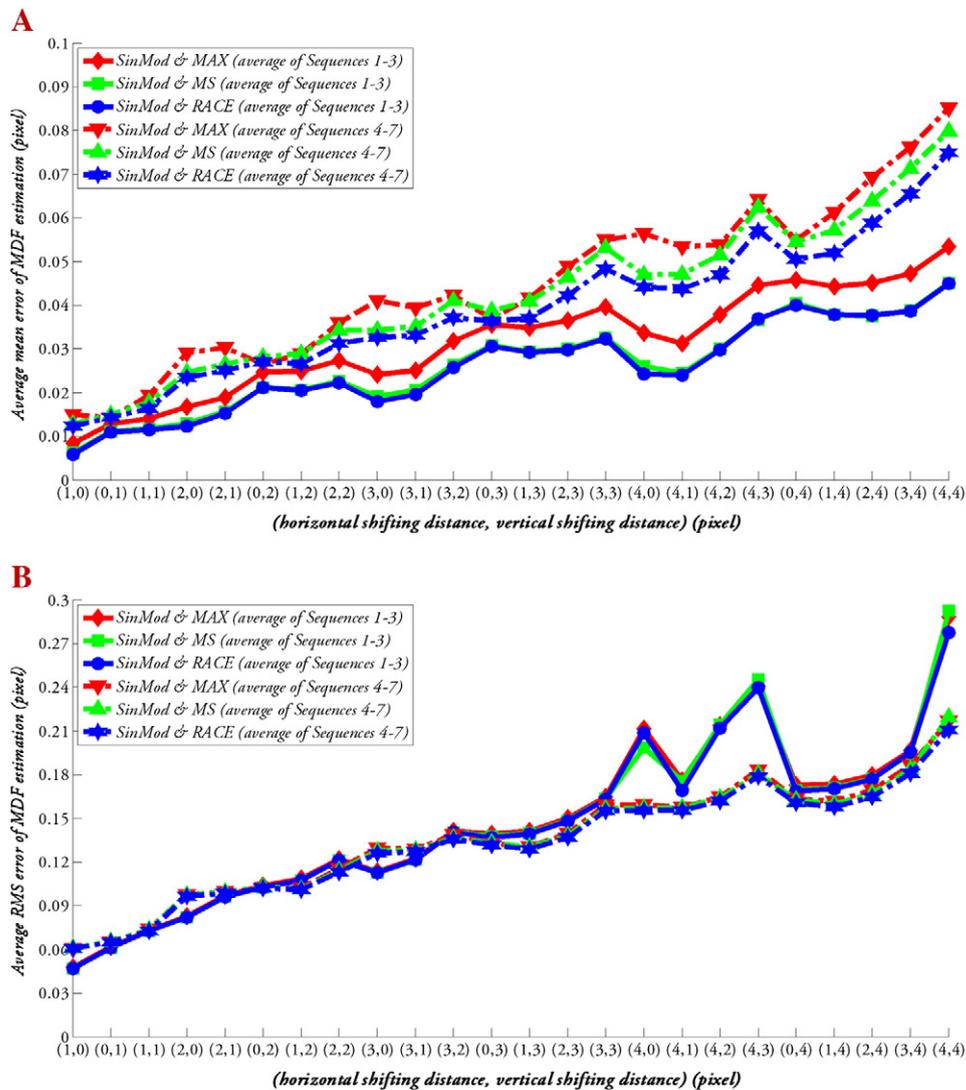


Fig. 11. A. Average mean error of MDF estimation vs. shifting distances for Sequences 1–3 and 4–7. B. Average RMS error of MDF estimation vs. shifting distances for Sequences 1–3 and 4–7.

5. Discussion

The HARP method is widely accepted because of its good performance, ease of implementation, and long history of investigation and application. As a kind of analogous method with relatively better performance, the SinMod method deserves better attention and promotion. In SinMod, a less noticed but still significant part is the explicit selection/estimation approach of the CF, which plays a very important role in the whole method. Throughout the related works of SinMod, we find none of them has expressly focused on this CF estimation problem. To make the SinMod method be thorough and more accurate, this study presents the RACE algorithm as a well solution for estimating the required CF automatically, robustly, accurately and quickly.

The RACE algorithm is based on the mean-shift algorithm and the two-direction-combination strategy. With a proper initialization, the mean-shift algorithm is able to get very close to the ground truth of CF with both high accuracy and fast speed. Some other algorithms may be able to achieve similar accuracies (e.g., the TRV algorithm with a very narrow traveling step), and some others may achieve even faster speed (e.g., the MAX algorithm), but as far as we know, the mean-shift algorithm is still optimal since it has the best comprehensive performance (see all the differences of results

between MS/RACE and MAX). As a key supplementary, the two-direction-combination strategy is employed to reduce the overall error caused by using only one wave-direction for CF estimation, or in other words, to enhance the accuracy and robustness of CF estimation. Its effectiveness is clearly validated all over all the experiments (see all the differences of results between RACE and MS). Furthermore, the strategy barely increases the time cost (see Table 1).

In the mean-shift algorithm, the initial shape-center, the region radius and the kernel function all have their own different effects on the CF estimation result. The initial shape-center directly decides which of the four candidate CFs (see Fig. 2) to be estimated, but it does not need to be set very accurately. As long as this parameter is chosen close to the spatial frequency of tags along one tag-direction, the mean-shift algorithm can always achieve an accurate CF estimate. Differently, the region radius should be watchfully set in a rational range, since smaller radius would bring more iterations and might possibly make the CF estimate to fall into an undesired local minimum, while larger radius could likely result in a totally wrong estimated location. About the kernel function, it is shown in our repetitive experiments that with different kernels, the performance of the RACE algorithm changes slightly in speed and accuracy within an acceptable range. In general, the parameter settings of our

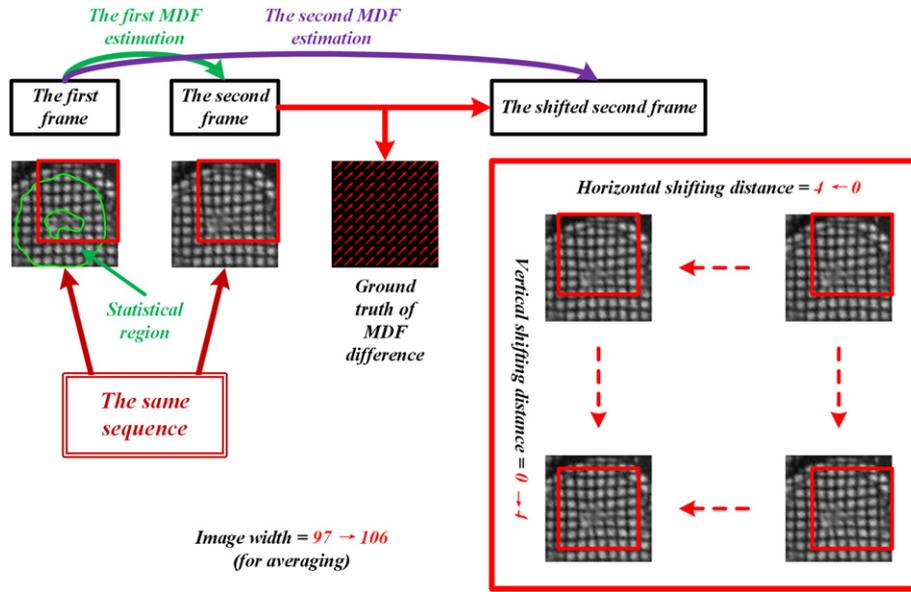


Fig. 12. The image selection scheme of the fourth experiment.

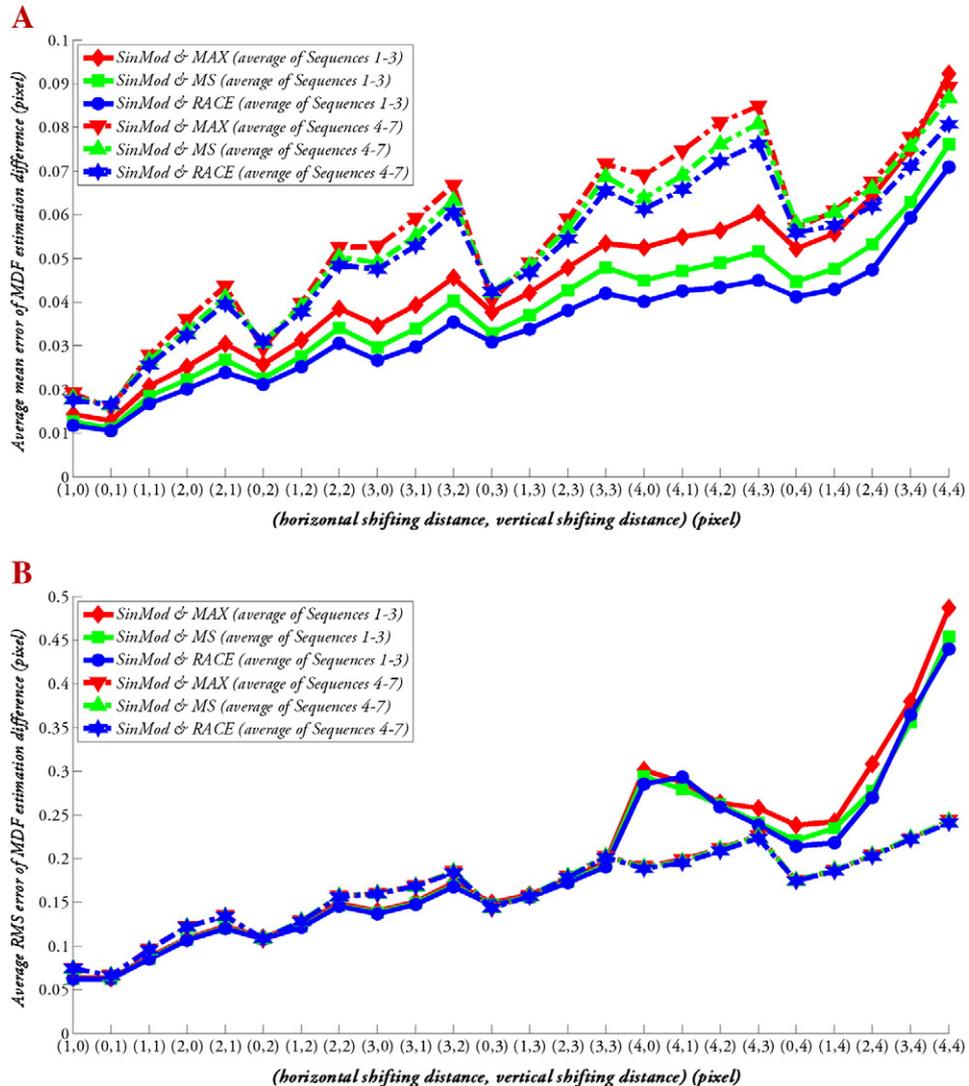


Fig. 13. A. Average mean error of MDF estimation difference vs. shifting distances for Sequences 1–3 and 4–7. B. Average RMS error of MDF estimation difference vs. shifting distances for Sequences 1–3 and 4–7.

Table 3

The relative error of each revised method.

| | Relative mean error (%) | | | Relative RMS error (%) | | |
|-----------------------------------|-------------------------|--------------------|---------------------------|------------------------|-------------|---------------------------|
| | SinMod & MAX | SinMod & MS | SinMod & RACE | SinMod & MAX | SinMod & MS | SinMod & RACE |
| Sequence 1 | 0.52 / 0.68 | 0.38 / 0.55 | <u>0.36</u> / <u>0.47</u> | 1.55 / 1.79 | 1.51 / 1.73 | <u>1.50</u> / <u>1.71</u> |
| Sequence 2 | 0.31 / 0.39 | 0.30 / 0.36 | <u>0.30</u> / <u>0.35</u> | 1.44 / 1.65 | 1.43 / 1.59 | 1.43 / <u>1.57</u> |
| Sequence 3 | 0.19 / 0.35 | <u>0.17</u> / 0.30 | 0.18 / <u>0.28</u> | 1.54 / 2.97 | 1.56 / 2.77 | 1.53 / <u>2.72</u> |
| Sequence 4 | 0.52 / <u>0.61</u> | 0.55 / 0.67 | <u>0.52</u> / 0.65 | 1.51 / <u>1.80</u> | 1.51 / 1.82 | <u>1.50</u> / 1.81 |
| Sequence 5 | 0.38 / <u>0.49</u> | 0.40 / 0.50 | <u>0.38</u> / 0.51 | 1.37 / <u>1.64</u> | 1.36 / 1.66 | <u>1.35</u> / 1.66 |
| Sequence 6 | 0.40 / 0.49 | 0.33 / 0.46 | <u>0.30</u> / <u>0.41</u> | 1.14 / 1.47 | 1.11 / 1.46 | <u>1.10</u> / <u>1.44</u> |
| Sequence 7 | 0.61 / 0.62 | 0.54 / 0.51 | <u>0.47</u> / <u>0.48</u> | 1.45 / 1.62 | 1.44 / 1.59 | <u>1.40</u> / <u>1.58</u> |
| Average of 1–3 (the volunteer) | 0.34 / 0.47 | 0.28 / 0.40 | <u>0.28</u> / <u>0.37</u> | 1.51 / 2.14 | 1.50 / 2.03 | <u>1.49</u> / <u>2.00</u> |
| Average of 4–7 (the patient) | 0.48 / 0.55 | 0.45 / 0.54 | <u>0.42</u> / <u>0.51</u> | 1.37 / 1.63 | 1.36 / 1.63 | <u>1.34</u> / <u>1.62</u> |

Data before '/' corresponds to the third experiment, while data after '/' corresponds to the fourth experiment. The boxes indicate the smallest error for each case.

current implementation may not be the most optimal, but the experiments show sufficient and effective results.

In conclusion, all the results in Subsections 4.4.1 and 4.4.2 together demonstrate that the RACE algorithm is definitely the most outstanding solution for the CF estimation problem under the circumstance that the specified tagging parameters are not known, as compared with all the other algorithms introduced in this paper. No matter on the robustness of CF estimation, or the accuracy of MDF estimation (also implies the accuracy of CF estimation), or the robustness of MDF estimation, RACE consistently achieves the best performance. Even in the aspect of processing time, RACE still performs very closely to the MAX algorithm which is supposed to be the fastest. By integrating with the RACE algorithm, the SinMod method would only have a very small deceleration, while the earnings are much larger.

6. Conclusions

A novel CF estimation problem for the SinMod method is studied and solved in this paper. Under the circumstances without known tagging parameters that can directly yield the ground truth of CF, three significant criterions which are all instructive in designing effective CF estimation algorithms are summarized from the original study of SinMod. Following these criterions, the RACE algorithm based on the mean-shift algorithm and the creative two-

direction-combination strategy is proposed to solve the CF estimation problem. Some other available CF estimation algorithms including MAX, TRV and MS are all provided for comparison. Two groups of human body in vivo TCMR data acquired from a healthy volunteer and a patient with cardiomyopathy are chosen for experiments. To address the no ground truth problem for real data, four validation approaches in total, including two for CF estimation and two for MDF estimation, are specially designed and applied to the four experiments, respectively. Several significant conclusions can be drawn from the experimental results: (1) the CF estimation problem is truly worth studying, since with different CF selection, the performance of SinMod alters between good and bad; (2) the two employed techniques in the proposed RACE algorithm, including the mean-shift algorithm and the two-direction-combination strategy, both have obvious and distinctive roles in facilitating the CF estimation; and (3) among all the valid CF estimation algorithms that we consider, the RACE algorithm has the most considerable comprehensive superiorities, as concretely embodied in high accuracy, strong robustness, and fast processing as well, of both CF and MDF estimation. With the application of RACE, people can study and carry out the SinMod method in a more convenient and effective way, and the follow-up studying and processing on cardiac motion and deformation analysis could be benefited to a certain degree.

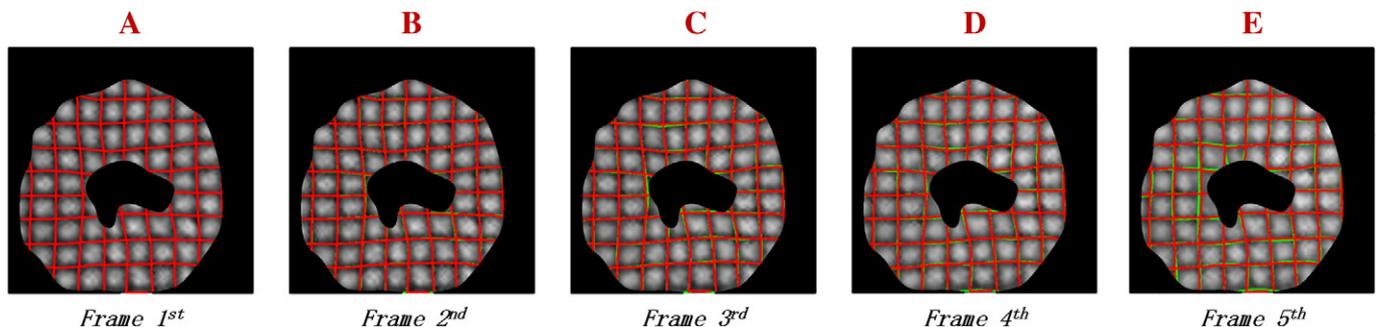


Fig. 14. A–E. Continuous tracking results of tags for the first five frames of Sequence 1, respectively.

References

- [1] Hendee WR, Morgan CJ. Magnetic resonance imaging. Part I—physical principles. *West J Med* 1984;141:491–500.
- [2] Wang H, Amini AA. Cardiac motion and deformation recovery from MRI: a review. *IEEE Trans Med Imaging* 2012;31:487–503.
- [3] Zerhouni EA, Parish DM, Rogers WJ, Yang A, Shapiro EP. Human heart: tagging with MR imaging—a method for noninvasive assessment of myocardial motion. *Radiology* 1988;169:59–63.
- [4] McVeigh ER. MRI of myocardial function: motion tracking techniques. *Magn Reson Imaging* 1996;14:137–50.
- [5] Kuijter JP, Marcus JT, Gotte MJ, van Rossum AC, Heethaar RM. Three-dimensional myocardial strain analysis based on short- and long-axis magnetic resonance tagged images using a 1-D displacement field. *Magn Reson Imaging* 2000;18:553–64.
- [6] Constantine G, Shan K, Flamm SD, Sivananthan MU. Role of MRI in clinical cardiology. *Lancet* 2004;363:2162–71.
- [7] Ibrahim ESH. Myocardial tagging by cardiovascular magnetic resonance: evolution of techniques—pulse sequences, analysis algorithms, and applications. *J Cardiovasc Magn Reson* 2011;13:36.
- [8] Osman NF, Kerwin WS, McVeigh ER, Prince JL. Cardiac motion tracking using CINE harmonic-phase (HARP) magnetic resonance imaging. *Magn Reson Med* 1999;42:1048–60.
- [9] Osman NF, McVeigh ER, Prince JL. Imaging heart motion using harmonic-phase MRI. *IEEE Trans Med Imaging* 2000;19:186–202.
- [10] Osman NF, Prince JL. Regenerating MR tagged images using harmonic-phase (HARP) methods. *IEEE Trans Biomed Eng* 2004;51:1428–33.
- [11] Khalifa AM, Youssef ABM, Osman NF. Improved harmonic-phase (HARP) method for motion tracking a tagged cardiac MR images. *Proc IEEE Eng Med Biol Soc* 2005:4298–301.
- [12] Pan L, Prince JL, Lima JAC, Osman NF. Fast tracking of cardiac motion using 3-D-HARP. *IEEE Trans Biomed Eng* 2005;52:1425–35.
- [13] Sampath S, Prince JL. Automatic 3-D tracking of cardiac material markers using slice-following and harmonic-phase MRI. *Magn Reson Imaging* 2007;25:197–208.
- [14] Liu X, Prince JL. Shortest path refinement for motion estimation from tagged MR images. *IEEE Trans Med Imaging* 2010;29:1560–72.
- [15] Arts T, Prinzen FW, Delhaas T, Milles JR, Rossi AC, Clarysse P. Mapping displacement and deformation of the heart with local sine-wave modeling. *IEEE Trans Med Imaging* 2010;29:1114–23.
- [16] Rossi AC, Arts T, Delhaas T. On the estimation of transmural myocardial shear by means of MRI tagging. *Proceedings of the 6th International Conference on Functional Imaging and Modeling of the Heart (FIMH)*, Lecture Notes in Computer Science (LNCC), 6666; 2011. p. 105–12.
- [17] Wang H, Amini AA. Accurate 2-D cardiac motion tracking using scattered data fitting incorporating phase information from MRI. *Proceedings of SPIE Medical Imaging 2010: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 76260 M; 2010 [7626:76260 M-1-12].
- [18] Wang H, Amini AA. Cardiac motion tracking with multilevel B-splines and SinMod from tagged MRI. *Proceedings of SPIE Medical Imaging 2011: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 796520; 2011 [7965:796520 -1-8].
- [19] Wang H, Amini AA. Cardiac deformation analysis using 3-D SinMod from 3-D CSPAMM tagged MRI. *Proceedings of SPIE Medical Imaging 2013: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 86720B; 2013 [8672:86720B-1-10].
- [20] Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans Inf Theory* 1975;21:32–40.
- [21] Cheng Y. Mean-shift, mode seeking, and clustering. *IEEE Trans Pattern Anal Mach Intell* 1995;17:790–9.
- [22] Lee S, Wolberg G, Shin SY. Scattered data interpolation with multilevel B-splines. *IEEE Trans Vis Comput Graph* 1997;3:228–44.
- [23] Axel L, Dougherty L. MR imaging of motion with spatial modulation of magnetization. *Radiology* 1989;171:841–5.
- [24] Axel L, Dougherty L. Heart wall motion: improved method of spatial modulation of magnetization for MR imaging. *Radiology* 1989;172:349–50.
- [25] Kusljevic MD. A simple recursive algorithm for frequency estimation. *IEEE Trans Instrum Meas* 2004;53:335–40.
- [26] Kusljevic MD. A simple recursive algorithm for simultaneous magnitude and frequency estimation. *IEEE Trans Instrum Meas* 2008;57:1207–14.
- [27] Park SY, Song YS, Kim HJ, Park J. Improved method for frequency estimation of sampled sinusoidal signals without iteration. *IEEE Trans Instrum Meas* 2011;60:2828–34.
- [28] Wang W, Xu Z, Lu W, Zhang X. Determination of the spread parameter in the Gaussian kernel for classification and regression. *Neurocomputing* 2003;55:643–63.
- [29] Liu X, Abd-Elmoniem KZ, Stone M, Murano EZ, Zhuo J, Gullapalli RP, et al. Incompressible deformation estimation algorithm (IDEA) from tagged MR images. *IEEE Trans Med Imaging* 2012;31:326–40.
- [30] Alessandrini M, Basarab A, Liebgott H, Bernard O. Myocardial motion estimation from medical images using the monogenic signal. *IEEE Trans Image Process* 2013;22:1084–95.
- [31] Chen T, Wang X, Chung S, Metaxas D, Axel L. Automated 3-D motion tracking using Gabor filter bank, robust point matching, and deformable models. *IEEE Trans Med Imaging* 2010;29:1–11.
- [32] Shi W, Zhuang X, Wang H, Duckett S, Luong DV, Tobon-Gomez C, et al. A comprehensive cardiac motion estimation framework using both untagged and 3-D tagged MR images based on nonrigid registration. *IEEE Trans Med Imaging* 2012;31:1263–75.
- [33] Kupinski MA, Hoppin JW, Clarkson E, Barrett HH, Kastis GA. Estimation in medical imaging without a gold standard. *Acad Radiol* 2002;9:290–7.