International Journal of Computer Assisted Radiology and Surgery (2019) 14:2165–2176
https://doi.org/10.1007/s11548-019-02030-z

ORIGINAL ARTICLE

Moving object tracking in clinical scenarios: application to cardiac surgery and cerebral aneurysm clipping

Sarada Prasad Dakua[1] · Julien Abinahed[1] · Ayman Zakaria[1] · Shidin Balakrishnan[1] · Georges Younes[1] · Nikhil Navkar[1] · Abdulla Al-Ansari[1] · Xiaojun Zhai[3] · Faycal Bensaali[2] · Abbes Amira[4]

Received: 28 January 2019 / Accepted: 3 July 2019 / Published online: 15 July 2019
© The Author(s) 2019

Abstract

Background and objectives Surgical procedures such as laparoscopic and robotic surgeries are popular since they are invasive in nature and use miniaturized surgical instruments for small incisions. Tracking of the instruments (graspers, needle drivers) and field of view from the stereoscopic camera during surgery could further help the surgeons to remain focussed and reduce the probability of committing any mistakes. Tracking is usually preferred in computerized video surveillance, traffic monitoring, military surveillance system, and vehicle navigation. Despite the numerous efforts over the last few years, object tracking still remains an open research problem, mainly due to motion blur, image noise, lack of image texture, and occlusion. Most of the existing object tracking methods are time-consuming and less accurate when the input video contains high volume of information and more number of instruments.

Methods This paper presents a variational framework to track the motion of moving objects in surgery videos. The key contributions are as follows: (1) A denoising method using stochastic resonance in maximal overlap discrete wavelet transform is proposed and (2) a robust energy functional based on Bhattacharyya coefficient to match the target region in the first frame of the input sequence with the subsequent frames using a similarity metric is developed. A modified affine transformation-based registration is used to estimate the motion of the features following an active contour-based segmentation method to converge the contour resulted from the registration process.

Results and conclusion The proposed method has been implemented on publicly available databases; the results are found satisfactory. Overlap index (OI) is used to evaluate the tracking performance, and the maximum OI is found to be 76% and 88% on private data and public data sequences.

Keywords Cerebral aneurysm · Segmentation · Object tracking · Heart surgery · Brain aneurysm clipping · Level sets

Introduction

Looking at the steep rise in cardiac diseases, bona fide treatment including surgery is necessary to prevent its rise and avoid sudden cardiac death [1]. Similarly, cerebral aneurysm (CA) is one of the devastating cerebrovascular diseases of adult population worldwide that cause subarachnoid hemorrhage, intracerebral hematoma, and other complications leading to high mortality rate [2]. Surgery is considered as an efficient modality for the patients with cardiac complications and ruptured cerebral aneurysms. Tracking could be considered as a treatment support and planning in robotic, laparoscopic, and medical education. During robotic surgery or laparoscopic surgery, the surgeons concentrate on the surgery to avoid even slight, possible mortality and morbidity and usually get stressed. In this scenario, motion tracking of the tools and viewing the desired operating field may be considered two supportive pillars to augment the treatment and improve success rate.

✉ Sarada Prasad Dakua
sdakua@hamad.qa

[1] Department of Surgery, Hamad Medical Corporation, Doha, Qatar

[2] Department of Electrical Engineering, Qatar University, Doha, Qatar

[3] School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK

[4] Faculty of Computing, Engineering and Media De Montfort University, Leicester, UK

Clinical requirements in surgery

Many factors contribute to successful outcome of a surgery, specifically minimally invasive surgery (MIS). These include technical factors, such as in-depth understanding of the relevant anatomy, clear understanding of the steps involved in the procedure, well-honed surgical skills and tool manipulation, as well as anthropomorphic factors such as operating team chemistry and dynamics. To a certain degree, MIS surgeons can advance their anatomy knowledge and procedural understanding through reading and surgical videos; however, other technical skills such as tool manipulation and positioning, which are very crucial to the successful outcome of the surgery [3,4] are more complex, nuanced and time dependent to develop due to restricted vision, limited working space, loss of visual cues and tactile feedback [5]. Quality and adequacy of surgical proficiency directly impact intra-operative and postoperative outcomes [6]. The existing "apprenticeship" model of training in surgery provides limited and time-consuming opportunities to gain the required technical competencies. In its current form, the assessment of surgical proficiency is heavily reliant on subject-matter experts/subjective assessments [3]. Thus, surgical training and planning could benefit greatly from visual support provided by instrument/motion tracking, by providing benchmarked metrics for continued objective and constructive assessment of highest standards of surgical skills, and lowering the risk of false tool trajectories and orientations [7], alignment of implants and placement of screws [8], etc.

Such augmented visual support for both surgical training and planning could be provided through object/motion tracking of the tools (such as scope, scissors, etc.) by providing objective assessment, benchmarking, and automated feedback on metrics such as path length/deviation, economy and smoothness of hand movements, depth perception, rotational orientation, changes in instrument velocity and time [9]. Zhao et al. [10] report that intra-operative tracking/detection of surgical instruments can provide important information to monitor instruments for the operational navigation in MIS, especially in the robotic minimally invasive surgeries (RMIS). Thus, based on the above, the perceived impact of tool tracking/positioning on surgical training and intra-operative guidance leads to (a) ensured patient safety via proficient tool movements and avoidance of critical tissue structures and (b) facilitation of a smooth and efficient invasive procedure [11]. This is crucial in surgery, as by continuously charting the location, movement, speed, and acceleration of the different surgical instruments in the operating field, the surgeon is continuously aware of the whereabouts of his instruments in relation to the patient's vital organs, blood vessels, and nerves during surgery. For surgical training, it objectively helps assess surgical performance and helps differentiate between an expert and a novice surgeon, such that optimal training can then be provided to the novice to ensure the highest levels of patient care [3]. Therefore, precise positioning of the tools remains pivotal in minimally invasive surgical procedures [12] highlighting the need of object tracking via its impact on surgical training and intra-operative guidance.

Kobayashi et al. [13] applied surgical navigation techniques and tool tracking to renal artery dissection within the robot-assisted partial nephrectomy procedure and found that inefficient tool movements involving "insert," "pull," and "rotate" motions, as well as time to visualize and dissect the artery were significantly improved owing to improved visualization and control over the tool and anatomy. Pediatric orthopedic surgeons found an increase in accuracy and a reduction in operating time when using image-guided surgical robotic systems to overcome the inaccuracies of hand-controlled tool positioning [14]; these robots achieve this by providing information about surgical tools or implants relative to a target organ (bone). In urology, motion tracking can greatly assist in outpatient procedures such as MRI and ultrasound-guided prostate biopsy, allowing the surgeon to accurately position and invade suspicious malignant zones for a tissue sample [15]. In interventional radiology, motion tracking can help track guide-wires during endovascular interventions and radiation therapy [16]. In addition to these, applications of surgical navigation systems and tool tracking/motion analysis are being explored in many other surgical fields, including ear-nose-and-throat (ENT) surgery [7], craniomaxillofacial surgery [17], cardiothoracic surgery [18], and orthopedic surgery [19].

Related work

The literature of motion tracking is rich; a few recent methods are included in this paper. Kim and Park [20] present a strategy that is based on edge information to assist object-based video coding, motion estimation, and motion compensation for MPEG 4 and MPEG 7 utilizing the human visual perception to provide edge information. However, the method critically depends on its ability to establish correct correspondences between points on the model edges and edge pixels in an image. Furthermore, this is a non-trivial problem especially in the presence of large inter-frame motions and cluttered environments. Subudhi et al. [21] propose a two-step method: spatio-temporal spatial segmentation and temporal segmentation that uses Markov random field (MRF) model and posteriori probability (MAP) estimation technique. Duffner and Garcia [22] present an algorithm for real-time single-object tracking, where a detector makes use of the generalized Hough transform with color and gradient descriptors; a probabilistic segmentation method is used for foreground and background color distributions. However, it is computationally expensive, especially when the number

of parameters is large. It also could be erroneous because the gradient information usually leads to error when noise level is high. Li et al. [23] suggest a method within the correlation framework (CF) that models a tracker maximizing the margin between the target and surrounding background by exploiting background information effectively. They propose to train a CF by multilevel scale supervision, which aims to make CF sensitive to the target scale variation. Then the two individual modules are integrated into one framework simplifying the tracking model. However, the computational load and efficiency are still two major concerns. Mahalingam et al. [24] propose a fuzzy morphological filter and blob detection-based method for object tracking. However, the performance gets deteriorated in the presence of noise, lack of illumination, and occlusion. Zhang et al. [25] propose a correlation particle filter (CPF) that combines a correlation filter and a particle filter. However, this tracker is still unable to deal with scale variation and partial occlusion. Yang et al. [26] present a method to analyze frames extracted from videos using kernelized correlation filters (KCF) and background subtraction (BS) (KCF-BS) to plot the 3D trajectory of cabbage butterfly. The KCF-BS algorithm is used to track the butterfly in video frames and obtain coordinates of the target centroid in two videos. However, it is noticed that the target sometimes gets lost and the method is unable to re-detect or recognize the target when the target motion is fast. Du et al. [27] propose an object tracking method for satellite videos by fusing KCF tracker and a three-frame difference algorithm. Although the method reports interesting results, it takes long time to perform. Liu et al. [29] propose a correlation filter-based tracker that consists of multiple positions' detections and alternate templates. The detection position is repositioned according to the estimated speed of target by an optical flow method, and the alternate template is stored with a template update mechanism. However, this method fails to perform if the size of each target is too small compared with the entire image, and the target and the background are very similar. Liu et al. [30] propose a method by integrating histogram of oriented gradient, RGB histogram, and motion histogram into a novel statistical model to track the target in unmanned aerial vehicle-captured videos. However, it fails to perform in occluded scenes.

Du et al. [31] present a method that is based on iterative graph seeking. Usually, the superpixel-based methods use mid-level visual cues to represent target parts where local appearance variations are exploited by superpixel representation. These methods have three sequential steps: (A) target part selection, (B) target part matching, and (C) target state estimation. (A) selects candidate target parts from the background, (B) a local appearance model associates parts between consecutive frames (target part matching, center pixel location and size of the target) is estimated based on majority voting, and (C) target state is estimated based on majority voting of matching results. This method integrates target part selection, part matching, and state estimation using a unified energy minimization framework. It incorporates structural information in local parts variations using the global constraint. Although the results are reported promising, the target part selection and target part matching when combinedly merge with the correlation filter, the estimation of the target takes long time to converge due to scale variation and partial occlusion that are bound to happen in surgery scenarios. Furthermore, when the noise level (for instance, in cardiac cine MRI data) in the input frames is high, the method would certainly struggle to perform. We intend to address these issues through our proposed method. Furthermore, if the literature above is carefully observed, noise has always been an issue in most of the methods. Therefore, in our proposed method, we first denoise the input frames. The target region on the first frame is chosen by a level set (LS) function, and then, the foreground and background models are generated. The foreground and background distributions are determined using the models in subsequent frames, and the motion of the pixels from the region of interest is estimated through a registration framework. Additionally, the selected region contour in the current frame is registered with the subsequent frame. Finally, segmentation is applied to refine the contour generated during registration and the contour is updated.

The paper is organized as follows: Section "Methodology and data" describes the denoising stage; "Target rendering" section presents the approach for target rendering (including region selection and developing models); "Registration" section defines a method for motion estimation through registration; "Segmentation" section presents the segmentation; "Results and discussion" section provides the results, while "Conclusions and future work" section concludes the paper.

Methodology and data

The method is illustrated in Fig. 1. First, the input frame is denoised to minimize the negative impact of noise on subsequent steps. The target region is then selected followed by the development of background models for motion estimation through a registration framework. Finally, the rough contour generated in registration step is further refined (by a proper segmentation method) and the contour is updated on subsequent frames.

Denoising of image sequences

Over the years, most of the methods address the noisy and cluttered medical images, mostly, by filtering that result significant degradation in image quality. One of the efficient approaches that counter noise and constructively utilize noise

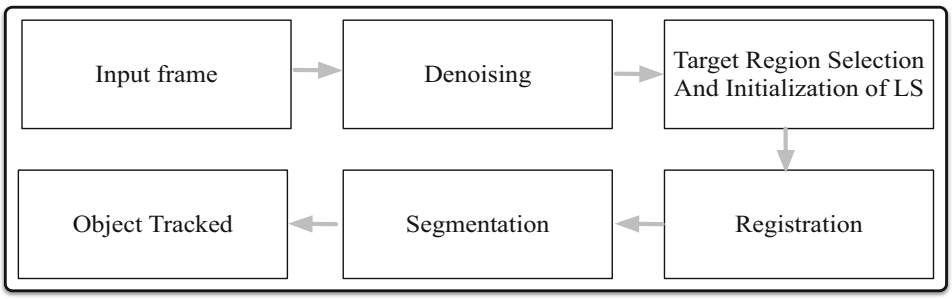

Fig. 1 Block diagram describing the proposed method

is stochastic resonance (SR) [33]. SR occurs if the signal-to-noise ratio (SNR) and input/output correlation have a well-marked maximum at a certain noise level. Unlike very low or high noise intensities, moderate ones allow the signal to cross the threshold giving maximum SNR at some optimum noise level. In the bistable SR model, upon addition of zero mean Gaussian noise, the pixel is transferred from weak signal state to strong signal state, which is modeled by Brownian motion of a particle (pc) placed in a double-well potential system. The state at which performance metrics are found optimum can be considered as the stable state providing maximum SNR. There have already been many attempts to use SR in different domains such as Fourier and spatial domains [34]; however, we have chosen the maximal overlap discrete wavelet transform (MODWT) [36] because of some of its key advantages: (1) MODWT can handle any sample size, (2) the smooth and detail coefficients of MODWT multiresolution analysis are associated with zero phase filters, (3) it is transform invariant, and (4) it produces a more asymptotically efficient wavelet variance estimator than DWT.

Maximal overlap discrete wavelet transform

Generally, DWT is defined by: $\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi\left(2^j t - k\right) \quad j, k \in \mathrm{Z};\ z = \{0, 1, 2, \ldots\}$, where ψ is a real-valued function compactly supported, and $\int_{-\infty}^{\infty}\psi(t)\,\mathrm{d}t = 0$. MODWT is evaluated using dilation equations: $\phi(t) = \sqrt{2}\sum_k l_k\phi(2t-k)$, $\psi(t) = \sqrt{2}\sum_k h_k\phi(2t-k)$, where $\phi(2t-k)$ and $\psi(t)$ are father wavelet defining low-pass filter coefficients and mother wavelet defining high-pass filter coefficients l_k: $l_k = \sqrt{2}\int_{-\infty}^{\infty}\phi(t)\,\phi(2t-k)\,\mathrm{d}t$, $h_k = \sqrt{2}\int_{-\infty}^{\infty}\psi(t)\,\psi(2t-k)\mathrm{d}t$.

Denoising by MODWT

In this methodology, 2D MODWT is applied to the $M \times N$ size image I. Applying SR to the approximation and detail coefficients, the stochastically enhanced (tuned) coefficient sets in MODWT domain are obtained as $W_{\psi}^{s}(l, p, q)_{SR}$ and $W(l_0, p, q)_{SR}$. The SR in discrete form is defined as: $\frac{\mathrm{dx}}{\mathrm{dt}} = \left[ax - ex^3\right] + B\sin\omega t + \sqrt{D}\xi(t)$, where $\sqrt{D}\xi(t)$ and $B\sin\omega t$ represent noise and input, respectively; these are replaced by MODWT sub-band coefficients. The noise term is the factor to produce SR; maximization of SNR occurs at the double-well parameter a. Implementation of SR on digital images necessitates the need for solving the stochastic differential equation using Euler–Maruyama's method [35] that gives the iterative discrete equation:

$$x(n+1) = x(n) + \Delta t\left[(ax(n) - ex^3(n)) + \mathrm{Input}(n)\right] \tag{1}$$

where a and e are the bistable parameters, whereas n and Δt represent iteration and sampling time, respectively. $Input$ denotes the sequence of input signal and noise, with the initial condition being $x(0) = 0$. The final stochastic simulation is obtained after some predefined number of iterations. Given the tuned (enhanced and stabilized) set of wavelet coefficients ($X_{\phi}(l_0, p, q)$ and $X_{\psi}^{s}(l, p, q)$), the denoised image I_{denoised} in spatial domain is obtained by inverse maximal overlap discrete wavelet transform (IMODWT) as:

$$I_{\mathrm{denoised}} = \frac{1}{\sqrt{MN}}\sum_{p}\sum_{q} X_{\phi}(l_0, p, q)\,\phi_{l_0,p,q}(i, j) + \frac{1}{\sqrt{MN}}\sum_{s\in(H,V,D)}\sum_{l=l_0}\sum_{p}\sum_{q} X_{\psi}^{s}(l, p, q)\,\psi_{l_0,p,q}^{s}(i, j)$$

The double-well parameters a and e are determined from the SNR by differentiating SR with respect to a and equating to zero; in this way, SNR is maximized resulting in $a = 2\sigma_0^2$ for maximum SNR, where σ_0 is the noise level administered to the input image. The maximum possible value of restoring force ($R = B\sin\omega t$) in terms of gradient of some bistable potential function $U(x)$, $R = -\frac{\mathrm{dU}}{\mathrm{dx}} = -ax + ex^3$, $\frac{\mathrm{dR}}{\mathrm{dx}} = -a + 3ex^2 = 0$ resulting in $x = \sqrt{a/3e}$. R at this value gives maximum force as $\sqrt{\frac{4a^3}{27e}}$ and $B\sin\omega t < \sqrt{\frac{4a^3}{27e}}$. Maximizing the left term (keeping $B = 1$), $e < \frac{4a^3}{27}$. In order to get the parameter values, we consider $a = w \times 2\sigma_0^2$, and $e = z \times \sqrt{\frac{4a^3}{27}}$; w and z are weight parameters for a and e. Initially, w is an experimentally chosen constant that later

becomes input image standard deviation dependent, while z is a number less than 1 to ensure sub-threshold condition of the signal. In this way, the noise in input image is countered and maximum information from the image is achieved.

Target rendering

Target region selection or target rendering [28,37] is the initial step in this motion tracking. Then the features (such as intensity, color, edge, texture, etc.) are selected that can appropriately describe the target. The notations used in target rendering are: fs—feature space, r—number of features, **fd**—foreground distribution (by the features), and **bd**—background distribution. The region is initialized on the first frame and represented by a level set function ϕ because of its flexibility in choosing the contour. The distributions of foreground ($\phi \geq 0$) and background ($-th < \phi < 0$, th is the threshold to restrict the region of interest into small area) regions are represented by $fg(\phi)$ and $bg(\phi)$, respectively, and match with **fd** and **bd**. Next, the foreground and background models are generated. Suppose the pixels $\{x_{f,i}\}_{i=1,\ldots,n_f}$ and $\{x_{b,i}\}_{i=1,\ldots,n_b}$ fall in foreground and background regions; the function $z : \Re^2 \rightarrow \{1,\ldots,r\}$ can be used to map the pixels (x_i) into the bin $b(x_i)$ in feature space. The probability of the feature space in the models is: $fd_{fs} = \frac{1}{n_f}\sum_{i=1}^{n_f}\delta\left[\left(x_{i,f}\right) - fs\right]$ and $bd_{fs} = \frac{1}{n_b}\sum_{i=1}^{n_b}\delta\left[\left(x_{i,f}\right) - fs\right]$, where δ is the Kronecker delta function and n_f and n_b are the number of pixels in foreground and background, respectively. The foreground and background distributions in the current frame candidate region ($-th < \phi < 0$) are obtained as:

$$fg(\phi) = \frac{1}{F_f}\sum_{i=1}^{n} H(\phi(x_i))\,\delta[b(x_i) - fs] \text{ and } bg(\phi) = \frac{1}{F_b}\sum_{i=1}^{n}(1 - H(\phi(x_i)))\,\delta[b(x_i) - fs] \quad (2)$$

$H(.)$ is the Heaviside function to select foreground region; F_f and F_b are the normalization factors.

Registration

Registration of the target in the first frame with the next subsequent frame is performed to estimate the affine deformation of the target. We determine the foreground and background distributions in the frames and match them with respective foreground and background models. We use Bhattacharyya metric [38] because it is computationally fast and is already being used in face recognition for years. Additionally, it has straightforward geometric interpretation. Since it is the cosine angle between **fd** and $fg(\phi)$ or between **bd** and $bg(\phi)$, higher value of the coefficient indicates better matching between candidate and target models. Thus, our similarity distance measure:

$$En_1(\phi) = \sum_{fs=1}^{r}\left(\sqrt{fg_{fs}(\phi)\,fd_{fs}} + \gamma\sqrt{bg_{fs}(\phi)\,bd_{fs}}\right) \quad (3)$$

where γ is the weight to balance the contribution from both foreground and background in the matching.

For deformation estimation, we have proposed a simple and efficient framework as follows. Suppose in the current frame, ϕ_0 is the target initial position and the contour is obtained by $\phi = 0$. The probabilities $fg(\phi_0) = \{fg_{fs}(\phi_0)\}_{fs=1,\ldots,r}$ and $bg(\phi_0) = \{bg_{fs}(\phi_0)\}_{fs=1,\ldots,r}$ are computed. Applying Taylor's expansion:

$$En_1(\phi) = \frac{1}{2}\left(\sum_{fs=1}^{r}\sqrt{fg_{fs}(\phi_0)\,fd_{fs}} + \sum_{fs=1}^{r} fg_{fs}(\phi)\sqrt{\frac{fd_{fs}}{fg_{fs}(\phi_0)}}\right) + \frac{1}{2}\gamma\left(\sum_{fs=1}^{r}\sqrt{bg_{fs}(\phi_0)\,bd_{fs}} + \sum_{fs=1}^{r} bg_{fs}(\phi)\sqrt{\frac{bd_{fs}}{bg_{fs}(\phi_0)}}\right) \quad (4)$$

By putting Eq. (2) in (4), we get:

$$En_1(\phi) = \frac{1}{2}\left(\sum_{fs=1}^{r}\sqrt{fg_{fs}(\phi_0)\,fd_{fs}} + \frac{1}{F_f}\sum_{fs=1}^{n} h_{f,i}H(\phi(x_i))\right) + \frac{1}{2}\gamma\left(\sum_{fs=1}^{r}\sqrt{bg_{fs}(\phi_0)\,bd_{fs}} + \frac{1}{F_b}\sum_{fs=1}^{n} h_{b,i}(1 - H(\phi(x_i)))\right) \quad (5)$$

where the weights that play a pivotal role in detecting the new centroid of the target are: $h_{f,i} = \sum_{fs=1}^{r}\sqrt{\frac{fd_{fs}}{fg_{fs}(\phi_0)}}\delta[z(x_i) - fs]$ *and* $h_{f,i} = \sum_{fs=1}^{r}\sqrt{\frac{bd_{fs}}{bg_{fs}(\phi_0)}}\delta[z(x_i) - fs]$. Higher value of Bhattacharyya coefficient can be obtained by maximizing (5) that is a function of location x and contour.

Furthermore, we consider the foreground and background intensity as additional feature. Suppose the first frame, $u_0(x, y)$, consists of two concentric regions (u_0^i, u_0^o) meaning the input image contains more than one intensity label. This is certainly challenging in determining a smooth contour initialization and deformation because of varying intensities. Therefore, we integrate both local and global image information in the energy term in order to make it perform as a perfect step detector with respect to the initialization of contour. The energy term is defined as:

$$En_2 = \lambda_1 E^G + \lambda_2 E^L + E^R \quad (6)$$

where λ_1 and λ_2 are fixed constants; E^G, E^L, and E^R are the global term, local term, and regularized term, respectively

(containing respective image information). E^R controls the boundary smoothness. The local term is defined as,

$$E^L = \int_{\phi<0} \frac{(g_k u_0(x,y) - u_0(x,y) - d_1(x,y))^2}{d_1(x,y)^2} \mathrm{d}x\mathrm{d}y + \int_{\phi>0} \frac{(g_k u_0(x,y) - u_0(x,y) - d_2(x,y))^2}{d_2(x,y)^2} \mathrm{d}x\mathrm{d}y \tag{7}$$

where g_k is an averaging filter with $k \times k$ size, d_1 and d_2 are intensity averages of the difference image $g_k u_0(x,y) - u_0(x,y)$ inside and outside the variable curve C, respectively. The global term:

$$E^G = \int_{\phi<0} \frac{(u_0(x,y) - c_1(x,y))^2}{c_1(x,y)^2} \mathrm{d}x\mathrm{d}y + \int_{\phi>0} \frac{(u_0(x,y) - c_2(x,y))^2}{c_2(x,y)^2} \mathrm{d}x\mathrm{d}y \tag{8}$$

where the constants c_1, c_2 represent the average intensity of $u_0(x,y)$ inside C and outside C, respectively. c_1 and c_2 are approximated by a weighted average of image intensity $u_0(p,q)$, where (p,q) is the neighborhood of (x,y). It means $c_1(x,y)$ and $c_2(x,y)$ are spatially varying; we formulate $c_1(x,y)$ and $c_2(x,y)$ as, $c_1(x,y) = \frac{\int_\Omega g_k((x,y)-(p,q)) u_0(p,q) H(\phi(p,q)) dpdq}{g_k((x,y)-(p,q)) H(\phi(p,q)) dpdq}$ and $c_2(x,y) = \frac{\int_\Omega g_k((x,y)-(p,q)) u_0(p,q)(1-H(\phi(p,q))) dpdq}{g_k((x,y)-(p,q))(1-H(\phi(p,q))) dpdq}$. We use the conventional regularizing term E_R that includes a penalty term on the total length of the edge contour for a given segmentation. Also it contains another penalty term on the total area of the foreground region found by the segmentation. The energy term therefore becomes:

$$\begin{aligned} En_2(\phi) = {} & \mu \int_\Omega \delta(\phi) + v \int_\Omega H(\phi(x,y)) \mathrm{d}x\mathrm{d}y + |\nabla\phi| \mathrm{d}x\mathrm{d}y \\ & + \lambda_1 \int_\Omega \frac{(u_0(x,y) - c_1(x,y))^2 H(\phi(x,y))}{c_1(x,y)^2} \mathrm{d}x\mathrm{d}y \\ & + \lambda_1 \int_\Omega \frac{(u_0(x,y) - c_1(x,y))^2 (1 - H(\phi(x,y)))}{c_2(x,y)^2} \mathrm{d}x\mathrm{d}y \\ & + \lambda_2 \frac{(g_k u_0(x,y) - d_1(x,y))^2 H(\phi(x,y))}{d_1(x,y)^2} \mathrm{d}x\mathrm{d}y \\ & + \lambda_2 \frac{(g_k u_0(x,y) - d_2(x,y))^2 (1 - H(\phi(x,y)))}{d_2(x,y)^2} \mathrm{d}x\mathrm{d}y \end{aligned} \tag{9}$$

This Eq. (9) has to be maximized to obtain higher Bhattacharyya coefficient. The similarity distance measure now becomes:

$$En(\phi) = En_1(\phi) + En_2(\phi) \tag{10}$$

We model the motion of the target as affine transformation by introducing a wrap in (10):

$$x = h(x, \Delta T) = \begin{pmatrix} 1+fg_1 & fg_3 & fg_5 \\ fg_2 & 1+fg_4 & fg_6 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{11}$$

The column vector characterizes the change in poses. Substituting (11) in (10) and omitting the terms that are not a function of ΔT-incremental warp (represented ϕ), we obtain:

$$\begin{aligned} En(\phi) = {} & \frac{1}{2F_f} \sum_{i=1}^{n} H(\phi(h(x, \Delta T))) w_{f,i} \\ & + \frac{1}{2F_b} \gamma \sum_{i=1}^{n} (1 - H(\phi(h(x, \Delta T)))) w_{b,i} \end{aligned} \tag{12}$$

ΔT tends to 0, and the estimation gets converged. In this way, the registration step iteratively estimates the shape change until it gets converged.

Segmentation

Since the tracker in the registration stage is still not able to extract the target contour properly, the registration result needs to be refined through segmentation. In order to do this, we optimize ϕ in Eq. (10) because the equation is a function of ϕ; in other words, $\frac{\partial En(\phi(x_i))}{\partial \phi(x_i)} = 0$. This is solved by well-known steepest-ascent method: $\frac{\partial En(\phi(x_i),t)}{\partial t} = \frac{\partial En(\phi(x_i))}{\partial \phi(x_i)}$. We obtain:

$$\begin{aligned} \frac{\delta\phi(x,y,t)}{\delta t} = {} & \delta_\in(\phi) \left[\mu \nabla \cdot \left(\frac{\nabla\phi}{|\nabla\phi|} \right) - v \right. \\ & \left. + \lambda_1 \left(\frac{(u_0(x,y) - c_2(x,y))^2}{c_2(x,y)^2} - \frac{(u_0(x,y) - c_1(x,y))^2}{c_1(x,y)^2} \right) \right] \\ & + \lambda_2 \left(\frac{(g_k u_0(x,y) - d_2(x,y))^2}{d_2(x,y)^2} - \frac{(g_k u_0(x,y) - d_1(x,y))^2}{d_1(x,y)^2} \right) \\ & + \frac{1}{2} \Delta t \delta_\in(\phi) \left(\frac{1}{F_f} h_{f,i} - \gamma \frac{1}{F_b} h_{b,i} \right) \end{aligned} \tag{13}$$

$$\frac{\partial_\varepsilon(\phi)}{|\nabla\phi|} \frac{\partial\phi}{\partial \vec{n}} = 0 \ on \ \partial\Omega \tag{14}$$

where H and $\delta_\in$ represent the Heaviside function and Dirac measure, respectively; $\frac{\partial\phi}{\partial \vec{n}}$ and $\vec{n}$ denote the normal derivative of ϕ at the boundary and the exterior normal to the boundary, respectively. Finally, the target is updated on subsequent frames.

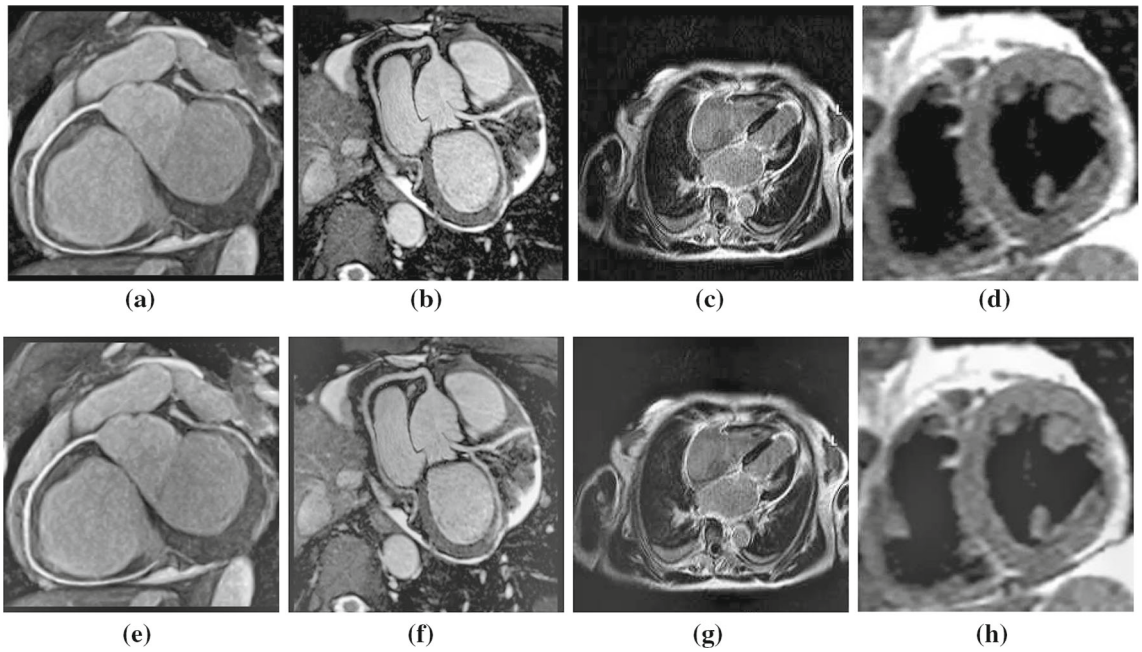

Fig. 2 **a**–**d** Input frames in a video sequence to be denoised. **e**,**f** Results of denoising

Data

The datasets used in this work are obtained from private sources such as Hamad Medical Corporation (30 data sequences) and public sources such as Sunnybrook [32] (45 data sequences) and VOT 2015 [40] (60 data sequences). The Sunnybrook Cardiac Data (SCD) consist of cine MRI data from a mixed set of patients and pathologies: healthy, hypertrophy, heart failure with infarction, and heart failure without infarction. Subset of this data set was first used in the automated myocardium segmentation challenge from short-axis MRI. The VOT 2015 sequences are chosen from a large pool of sequences including ALOV, OTB, non-tracking, Computer Vision Online, Professor Bob Fisher's Image Database, Videezy, Center for Research in Computer Vision, University of Central Florida, USA, NYU Center for Genomics and Systems Biology, Data Wrangling, Open Access Directory and Learning and Recognition in Vision Group, INRIA. The initial pool of sequences is created by combining the sequences from all the sources. After removal of duplicate sequences, grayscale sequences and sequences that contained objects with area smaller than 400 pixels, the final sequences are obtained; more details can be obtained from the Web site [45].

Results and discussion

Results

The proposed method is implemented on both private and public databases as described earlier. The qualitative results of denoising are provided in Fig. 2. We have quantitatively compared the proposed denoising method with that of Fourier because of its huge popularity [34]. The perceptual quality measurement (PQM) [41] is provided in Fig. 3, which shows greater value in case of MODWT suggesting higher efficacy of MODWT; in this figure, m denotes the mass of the particle that moves under stochastic condition. For denoising of the input images, the initial values of Δt and z are taken as 0.007 and 0.000027, respectively. To determine the quality of the denoised image, we have calculated distribution separation measure that estimates the degree of image quality. The DSM is defined as [34]: $\mathrm{DSM} = \left|\mu_T^E - \mu_B^E\right| - \left|\mu_T^O - \mu_B^O\right|$, where μ_T^E and μ_T^O are the mean of the selected target regions of the denoised and original images, respectively; μ_B^E and μ_B^O are the mean of the selected background region of the denoised and original image, respectively. The higher the value of DSM, the better is the quality. It is observed that the value of DSM is maximum at iteration 200 and then it starts decreasing; therefore, this iteration is considered as the optimal.

These denoised frames are further used in the subsequent steps in the proposed method. As mentioned earlier, we have included the image sequences of cardiac surgery and clipping for ruptured cerebral aneurysms in this work. We have also tested our method on cardiac cine MRI datasets, high contrast and low contrast levels, to highlight the performing capability of the method in varying intensities. The performance results on these datasets are provided in Figs. 4 and 5. We have chosen different scenarios for cerebral aneurysm surgical procedure (clipping): One is to track the scissors'

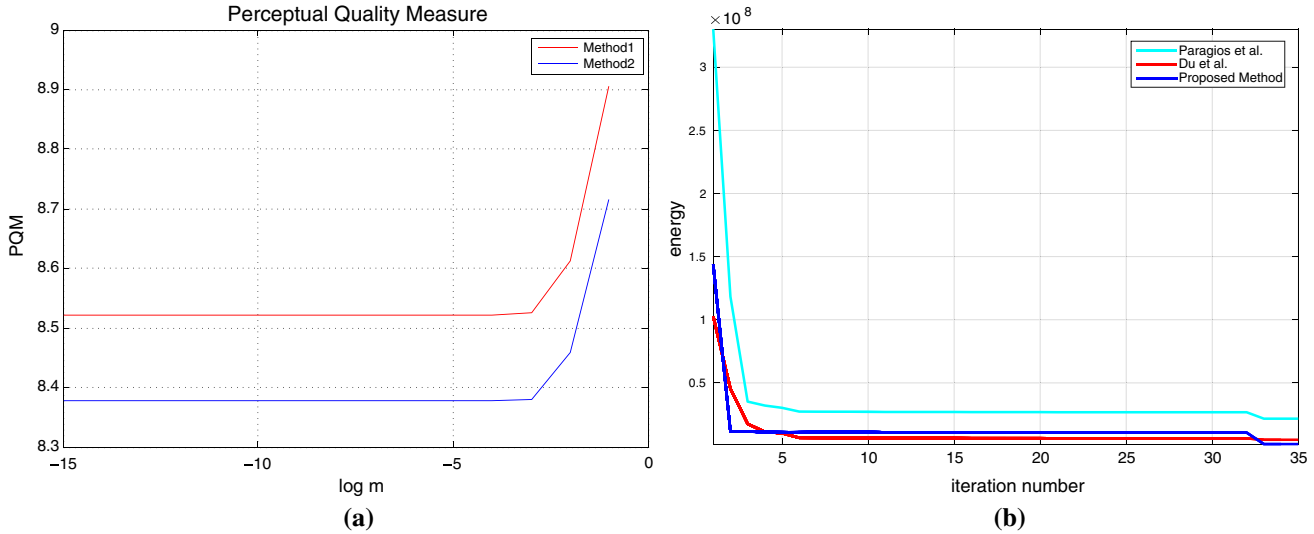

Fig. 3 Perceptual quality measures by Fourier (Method 2) and MODWT (Method 1); *m* in x-axis denotes the mass of the particle that moves under stochastic condition. **b** Energy convergence comparison of three methods

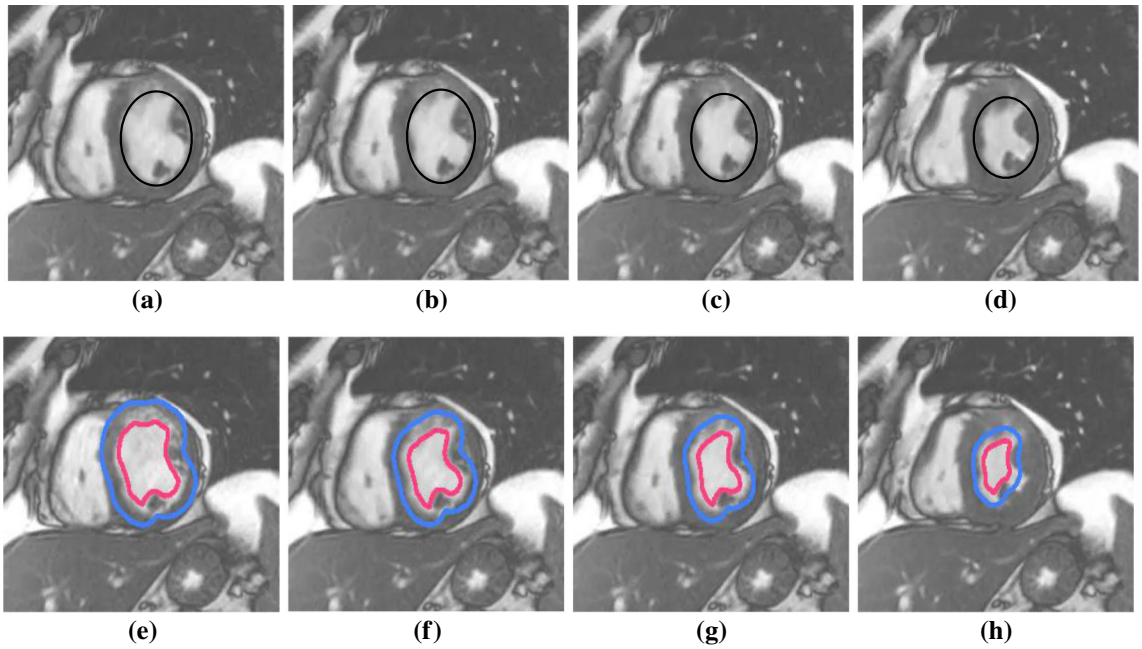

Fig. 4 **a**–**d** Ground truth frames. **e**–**h** Tracking of left ventricle in low-contrast cine magnetic resonance imaging (low-contrast CMRI) during cardiac surgery

or clippers' movement and the other one is to focus on the operating field during surgery, where multiple tools are used by the surgeons. It is important to track the motion of the scissors in order to minimize the damage caused by their movement. Besides the tools' tracking, capturing or tracking the operating field is also important; it helps the surgeon in concentrating on the tools used during the surgery and the impacted tissues of interest. The results are given in Figs. 6 and 7. We have also tested the proposed method on VOT 2015 datasets and found some satisfactory results as can be observed in Fig. 8. We have included this particular dataset in this paper to emphasize on the fact that the foreground is not very significantly different than the background like it happens in medical data sequences. Usually, the medical data are blurry (either reddish or grayish) and lack contrast as can be observed from the figures. In this scenario, only a contour surrounding the tools could easily be ignored; therefore, just for user's (surgeon) convenience, we have added the blue line surrounding the red line in the tracking results. While calculating the accuracy, red line is only taken into consideration. In order to determine the segmentation accuracy, we have used Dice coefficient (DC), which may be defined as

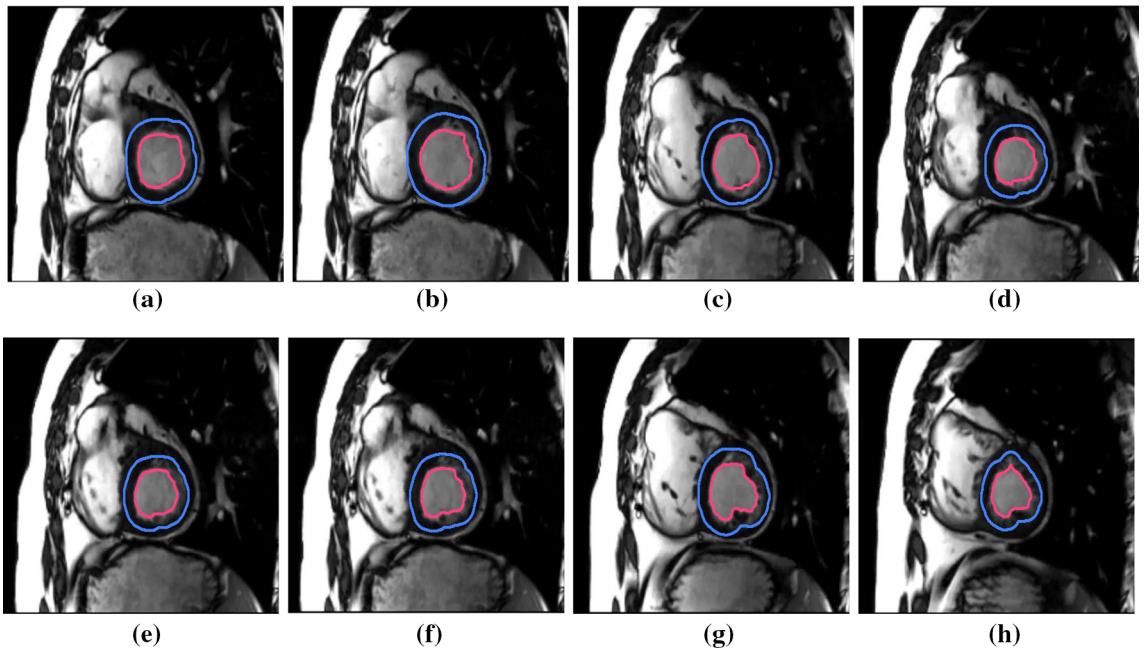

Fig. 5 Tracking of left ventricle in high-contrast cine magnetic resonance imaging (high-contrast CMRI) during cardiac surgery

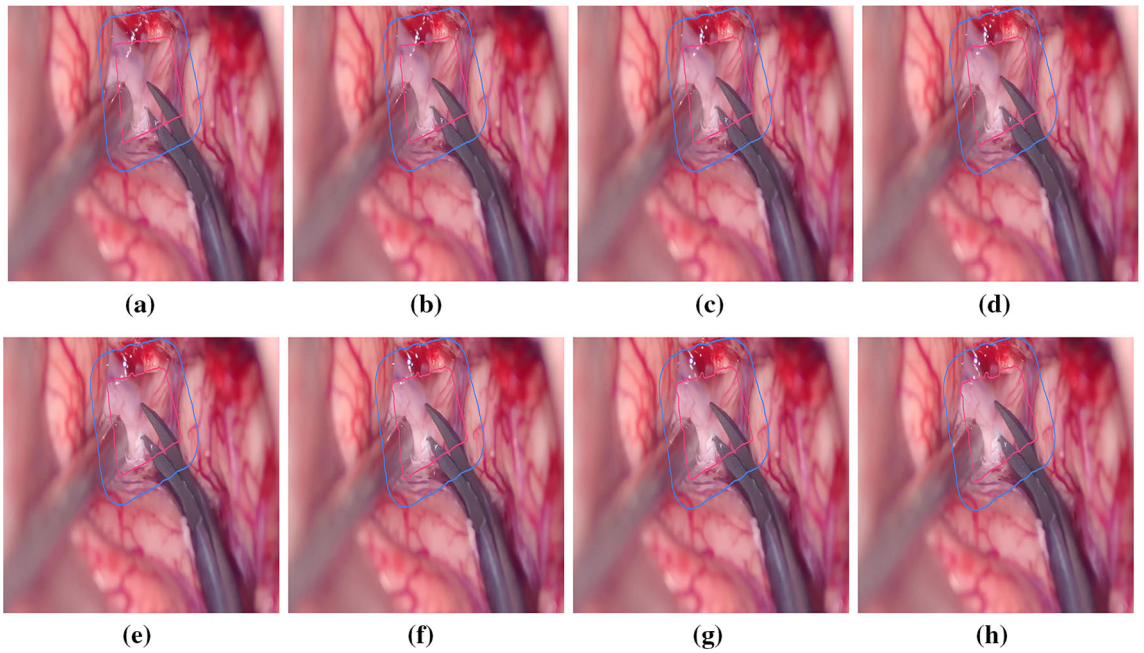

Fig. 6 Tracking of the operating field with multiple objects during cerebral aneurysm clipping

[44]: $DC = 2 \times \frac{|X \cap Y|}{|X|+|Y|}$, where X and Y are two point sets. The average segmentation accuracy on 3-T machine is 94%, whereas in case of 7 T, it is found to be 96%. The proposed method has performed as expected, which can be verified from the results provided in "Results" section. We have optimized the algorithm and code; average time taken to perform tracking and average number of frames are less than 25–30 s and 24 frames per second, respectively. We have also compared the performance of the proposed method with other similar methods ([31,42]); the proposed method converges faster than the other methods 3(b). We have also calculated overlap index (OI) [43] to determine the overlap between the resulting target contour and the actual boundary. We have found it highest in case of the proposed method against others as can be observed from Table 1.

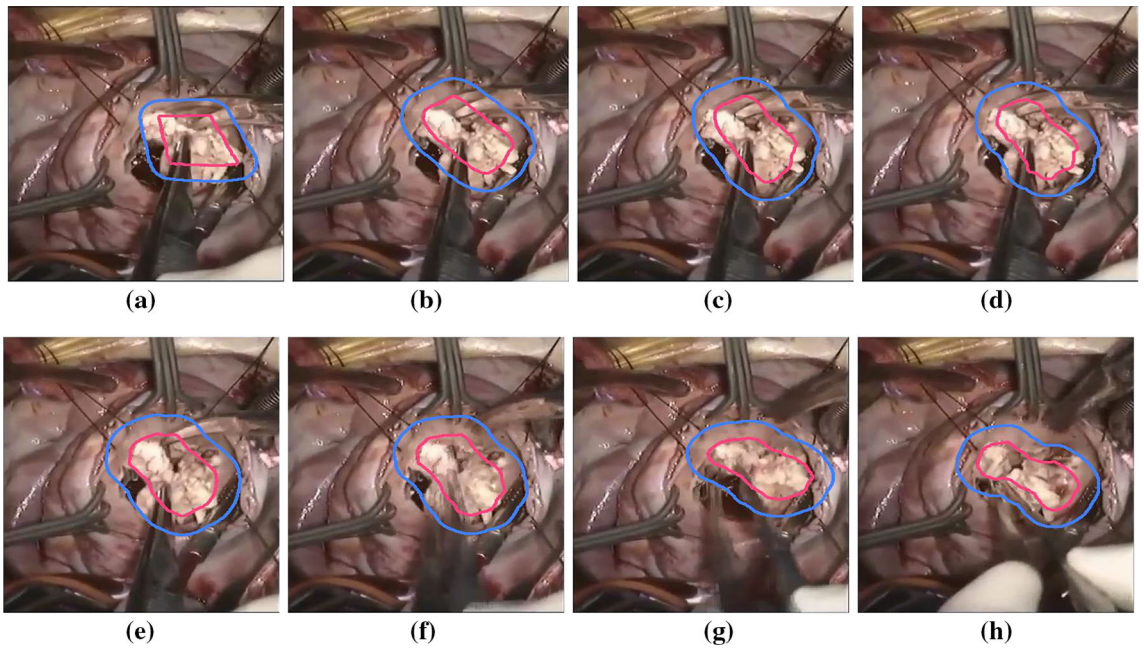

Fig. 7 Tracking of the operating field with multiple objects during cardiac surgery

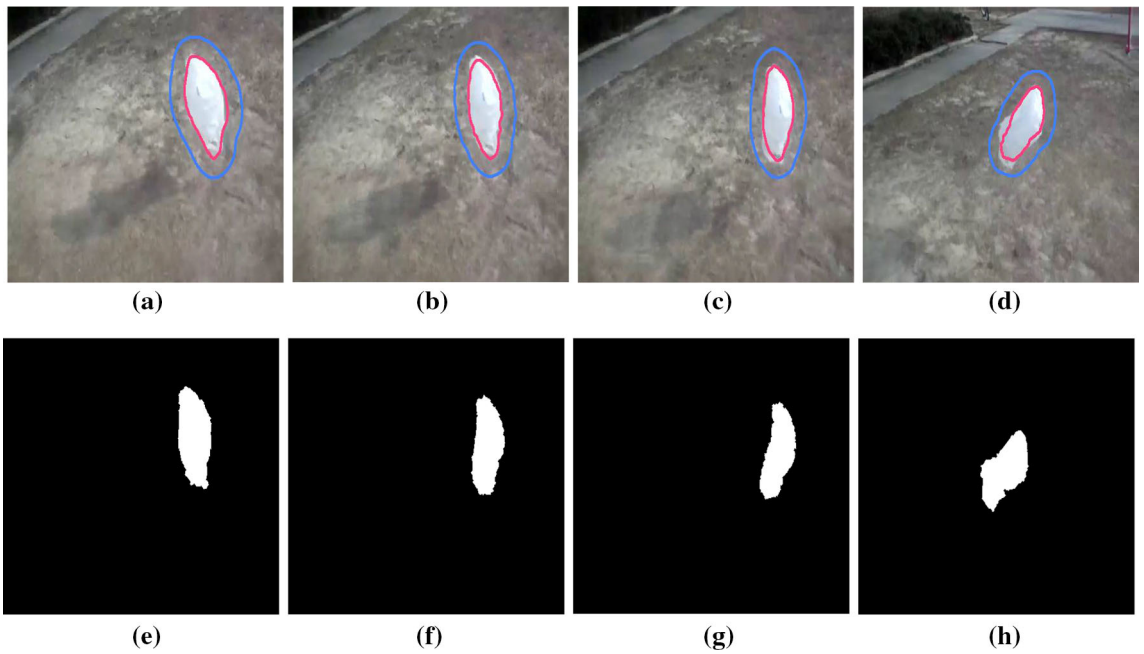

Fig. 8 **a–d** Tracked frames in a video sequence (VOT 2015). **e–f** Corresponding ground truth sequences

Discussion

The values of bistable system parameters play a crucial role in the process of denoising using SR. The expression for SR on any data set contains additive terms of multiples of w and subtractive term of multiples of z. This is observed that the images that have low contrast and low dynamic range require larger values of w, while those that have relatively more contrast and cover an appreciable gray level range require smaller values of w for proper denoising. Values of Δt have been studied to be similar to that of w. This is also perceived that w is inversely proportional to overall variance signifying the contrast of input image. Optimization process leads us to the optimum value of w; the value of z should be less than 1 so that condition $e < \sqrt{\frac{4a^3}{27}}$ holds assuring that the system is bistable and signal is sub-threshold so that SR can be appli-

Table 1 Overlap index comparison of different methods on hospital and VOT 2015 datasets

Method	Hospital and SCD datasets			
	FOV-CA (%)	Scissors-CA (%)	Low-contrast CMRI (%)	High-contrast CMRI (%)
Paragois et al. [42]	64	65	64	65
Du et al. [31]	67	69	68	72
Proposed method	**72**	**74**	**73**	**76**
Method	VOT 2015 Dataset			
	Rabit (%)	Shaking (%)	Racing (%)	Octopus(%)
Paragois et al. [42]	69	70	68	71
Du et al. [31]	72	75	70	76
Proposed method	**83**	**85**	**82**	**88**

The results of the proposed methods are shown in bold

cable. We prefer a very small value of this factor to remain well within the allowable range of e. Finally, we have noticed that the varying segmentation accuracy depends on the quality of the input data sequence. The MRI data obtained from 7-T machine give better accuracy than 3-T MRI machine.

Conclusions and future work

A variational framework has been presented to track the motion of moving objects and field of view in surgery sequences. We have presented a method that has used SR to denoise the input frames and a combined registration–segmentation framework to conduct motion tracking. We have introduced a robust similarity metric and an efficient energy functional in this framework. Despite the fact that the input data contain varying illumination, motion blur, lack of image texture, occlusion, and fast object movements, the performance of the proposed method is found quite satisfactory. In future, we intend to extensively evaluate the method quantitatively so that it can be well tested before trying in clinical practice.

Acknowledgements Open Access funding provided by the Qatar National Library. This work was partly supported by NPRP Grant #NPRP 5-792-2-328 from the Qatar National Research Fund (a member of Qatar Foundation).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Di Salvo TG, Acker MA, Dec GW, Byrne JG (2010) Mitral valve surgery in advanced heart failure. J Am Coll Cardiol 55:271–282
2. Zhai X, Eslami M, Hussein ES, Filali MS, Shalaby ST, Amira A, Bensaali F, Dakua S, Abinahed J, Al-Ansari A, Ahmed AZ (2018) Real-time automated image segmentation technique for cerebral aneurysm on reconfigurable system on chip. J Comput Sci 27:35–45
3. Ganni S, Botden SMBI, Chmarra M (2018) A software-based tool for video motion tracking in the surgical skills assessment landscape. Surg Endosc 32:2994
4. Jakimowicz JJ, Buzink S (2015) Training curriculum in minimal access surgery. In: Francis N, Fingerhut A, Bergamaschi R, Motson R (eds) Training in minimal access surgery. Springer, London, pp 15–34
5. Feng C, Haniffa H, Rozenblit JW, Peng J, Hamilton AJ, Salkini M (2006) Surgical training and performance assessment using a motion tracking system. In: International mediterranean modelling multiconference, I3M, pp 647–652
6. Carroll SM, Kennedy AM, Traynor O, Gallagher AG (2009) Objective assessment of surgical performance and its impact on a national selection programme of candidates for higher surgical training in plastic surgery. J Plast Reconstr Aesthet Surg 62:1543–1549
7. Pruliere-Escabasse V, Coste A (2010) Image-guided sinus surgery. Eur Ann Otorhinolaryngol Head Neck Dis 127:33–39
8. Tjardes T, Shafizadeh S, Rixen D, Paffrath T, Bouillon B, Steinhausen ES, Baethis H (2010) Image-guided spine surgery: state of the art and future directions. Eur Spine J 19:25–45
9. Shaharan S, Nugent E, Ryan DM, Traynor O, Neary P, Buckley D (2016) Basic surgical skill retention: can patriot motion tracking system provide an objective measurement for it? J Surg Educ 73:245–9
10. Zhao Z, Voros S, Weng Y, Chang F, Li R (2017) Tracking-by-detection of surgical instruments in minimally invasive surgery via the convolutional neural network deep learning-based method. Comput Assist Surg 22:26–35
11. Zhang M, Wu B, Ye C, Wang Y, Duan J, Zhang X, Zhang N (2019) Multiple instruments motion trajectory tracking in optical surgical navigation. Opt Express 27:15827–15845
12. Berry D (2009) Percutaneous aortic valve replacement: an important advance in cardiology. Eur Heart J 30:2167–2169
13. Kobayashi S, Cho B, Huaulme A, Tatsugami K, Honda H, Jannin P, Hashizumea M, Eto M (2019) Assessment of surgical skills by using surgical navigation in robot-assisted partial nephrectomy. Int

J Comput Assist Radiol Surg. https://doi.org/10.1007/s11548-019-01980-8
14. Docquier PL, Paul L, TranDuy K (2016) Surgical navigation in paediatric orthopaedics. EFORT Open Rev 1:152–159
15. Tadayyon H, Lasso A, Kaushal A, Guion P, Fichtinger G (2011) Target motion tracking in MRI-guided transrectal robotic prostate biopsy. IEEE Trans Biomed Eng 58:3135–42
16. Ozkan E, Tanner C, Kastelic M, Mattausch O, Makhinya M, Goksel O (2017) Robust motion tracking in liver from 2D ultrasound images using supporters. Int J Comput Assist Radiol Surg 12:941–950
17. Liu TJ, Ko AT, Tang YB, Lai HS, Chien HF, Hsieh TM (2016) Clinical application of different surgical navigation systems in complex craniomaxillofacial surgery: the use of multisurface 3-dimensional images and a 2-plane reference system. Ann Plast Surg 76:411–9
18. Engelhardt S, Simone RD, Al-Maisary S, Kolb S, Karck M, Meinzer HP, Wolf I (2016) Accuracy evaluation of a mitral valve surgery assistance system based on optical tracking. Int J Comput Assist Radiol Surg 11:1891–904
19. Niehaus R, Schilter D, Fornaciari P, Weinand C, Boyd M, Ziswiler M, Ehrendorfer S (2017) Experience of total knee arthroplasty using a novel navigation system within the surgical field. Knee 24:518–524
20. Kim BG, Park DJ (2004) Unsupervised video object segmentation and tracking based on new edge features. Pattern Recognit Lett 25:1731–1742
21. Subudhi BN, Nanda PK, Ghosh A (2011) A change information based fast algorithm for video object detection and tracking. IEEE Trans Circuits Syst Video Technol 21:993–1004
22. Duffner S, Garcia C (2017) Fast pixel wise adaptive visual tracking of non rigid objects. IEEE Trans Image Process 26:2368–2380
23. Li J, Zhou X, Chan S, Chen S (2017) Robust object tracking via large margin and scale adaptive correlation filter. IEEE Access 6:12642–12655
24. Mahalingam T, Subramoniam M (2018) A robust single and multiple moving object detection, tracking and classification. Appl Comput Inf. https://doi.org/10.1016/j.aci.2018.01.001
25. Zhang T, Liu S, Xu C, Liu B, Yang M (2018) Correlation particle filter for visual tracking. IEEE Trans Image Process 27:2676–2687
26. Yang yang G, Dong-jian H, Cong L (2018) Target tracking and 3D trajectory acquisition of cabbage butterfly based on the KCF-BS algorithm. Sci Rep Nat 8:9622
27. Du B, Sun Y, Cai S, Wu C, Du Q (2018) Object tracking in satellite videos by fusing the kernel correlation filter and the three-frame-difference algorithm. IEEE Trans Image Process 15:168–1821
28. Ning J, Zhang L, Zhang D, Yu W (2013) Joint registration and active contour segmentation for object tracking. IEEE Trans Circuits Syst Video Technol 23:1589–1597
29. Liu G, Liu S, Muhammad K, Sangaiah A, Doctor F (2018) Object tracking in vary lighting conditions for fog based intelligent surveillance of public spaces. IEEE Access 6:29283–29296
30. Liu S, Feng Y (2018) Real-time fast moving object tracking in severely degraded videos captured by unmanned aerial vehicle. Int J Adv Robot Syst SAGE 11:1–10
31. Du D, Wen L, Qi H, Huang Q, Tian Q, Lyu S (2018) Iterative graph seeking for object tracking. IEEE Trans Image Process 27:1809–1821
32. Radau P, Lu Y, Connelly K, Paul G, Dick AJ, Wright GA (2009) Evaluation framework for algorithms segmenting short axis cardiac MRI. MIDAS J Cardiac MR Left Ventricle Segm Chall
33. Dakua S, Abinahed J, Al-Ansari A (2018) A PCA based approach for brain aneurysm segmentation. J Multi Dimens Syst Signal Process 29:257–277
34. Rallabandi V, Roy P (2010) MRI enhancement using stochastic resonance in Fourier domain. Magn Reson Imaging 28:1361–1373
35. vom Scheidt J, Gard TC Introduction to stochastic differential equations. Pure and applied mathematics 114, XI, 234 pp. Marcel Dekker Inc., New York . ISBN 0-8247-7776-X
36. Yao SJ, Song YH, Zhang LZ, Cheng XY (2000) MODWT and networks for short-term electrical load forecasting. Energy Convers Manag 41:1975–1988
37. Comaniciu D, Ramesh V, Meer P (2003) Kernel-based object tracking. IEEE Trans Pattern Anal Mach Intell 25:564–577
38. Vezzetti E, Marcolin F (2015) Similarity measures for face recognition. Bentham Books, Sharjah, United Arab Emirates. ISBN: 978-1-68108-045-1
39. Erdem CE, Sankur B, Tekalp AM (2004) Performance measures for video object segmentation and tracking. IEEE Trans Image Process 13:937–951
40. Matej K, Matas J, Leonardis A, Felsberg M, Cehovin L (2015) The visual object tracking VOT 2015 challenge results. In: IEEE international conference on computer vision workshop, pp 564–586
41. Wang Z, Sheikh HR, Bovik AC (2002) No-reference perceptual quality assessment of jpeg compressed images. In: Proceedings of IEEE international conference image processing, page 477–480
42. Paragios N, Deriche R (2000) Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Trans Pattern Anal Mach Intell 22:266–280
43. Rosenfield GH, Fitzpatrick Lins K (1986) A coefficient of agreement as a measure of thematic classification accuracy. Photogramm Eng Remote Sens 52:223–227
44. Shi F, Yang Q, Guo X, Qureshi T, Tian Z, Miao H, Dey D, Li D, Fan Z (2019) Vessel wall segmentation using convolutional neural networks. IEEE Trans Biomed Eng. https://doi.org/10.1109/TBME.2019.2896972
45. http://www.votchallenge.net/vot2015/dataset.html

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.