Tumor Motion Tracking in Liver Ultrasound Images Using Mean Shift and Active Contour

Jalil Rasekhi

Department of Biomedical Engineering, Noshirvani University of Technology, Babol, Iran
jalil.rasekhi@gmail.com

Abstract

In this paper we present a new method for motion tracking of tumors in liver ultrasound image sequences. Our algorithm has two main steps. In the first step, we apply mean shift algorithm with multiple features to estimate the center of the target in each frame. Target in the first frame is defined using an ellipse. Edge, texture, and intensity features are extracted from the first frame, and then mean shift algorithm is applied to each feature separately to find the center of ellipse related to that feature in the next frame. The center of ellipse will be the weighted average of these centers. By using mean shift actually we estimate the target movement between two consecutive frames. Once the correct ellipsoid in each frame is known, in the second step we apply the Dynamic Directional Gradient Vector Flow (DDGVF) version of active contour models, in order to find the correct boundary of tumors. We sample a few points on the boundary of active contour then translate those points based on the translation of the center of ellipsoid in two consecutive frames to determine the target movement. We use these translated sample points as an initial guess for active contour in the next frame. Our experimental results show that, the suggested method provides a reliable performance for liver tumor tracking in ultrasound image sequences.

Keywords: Ultrasound image; liver cancer; tumor tracking; mean shift; active contour

1. Introduction

Liver cancer is one the most common types of cancer worldwide, and each year about 700000 people are diagnosed with this cancer, and which is responsible for 600000 deaths annually. In 2014 more than 33000 were diagnosed with liver cancer in USA only, resulting in 23000 deaths. Currently several medications are provided for the liver cancer. These include surgery, ablation through heating by microwaves or radio waves, radiotherapy, and chemotherapy. Radiotherapy is an effective invasive method for those patients suffering from a few number of small tumors, where doctors rule out the unjustifiable risky surgeries. In this method, doctors use high-energy radioactive rays either to eliminate cancer cells or to reduce the size of tumor (Goyal et al., 2014). During radiotherapy treatment, the inevitable breathing of the patients can displace tumors, which may hurt the
neighboring healthy tissues (Sharp et al., 2004). To reduce the side effects of radiation therapy and to protect the healthy tissues, the high-energy ray should be exclusively focused on the tumor. Therefore, the real-time estimation of the tumor’s location, and the adjustment of the ray focus during this process would significantly reduce from the hazardous side effects (Shirato et al., 2000, Shimizu et al., 2001, Buzurovic et al., 2010, Rubin et al., 2012, Cifor et al., 2013, Petrusca et al., 2013).

In recent years ultrasound images (US) have emerged as important resources for the medical community. The lower cost of US images compared with the computed tomography (CT) scan and magnetic resonance imaging (MRI) makes them a good and affordable tool for most doctors in studying their patients. Although the quality of US images is low compared to CT and MRI images, however these monochrome intensity-based images are capable of providing us with real time image sequences, which make them very attractive for applications such as tracking (Preiswerk et al., 2012, Rubin et al., 2012, Youngkyoo et al., 2012). Furthermore, US imaging has an excellent safety record. Despite X-rays or other types of imaging systems that use ionizing radiation, US is based on non-ionizing radiation, so that the patient will not face the same risks.

In this paper we approach the problem of tumor tracking in US image sequences for a specific patient suffering from liver cancer. Our tracking method is composed of two steps. In the first step, we use mean shift algorithm (Comaniciu et al., 2002) with multiple features to estimate the center of the ellipse in the next frame. We use intensity, texture and edge histograms as the features. Similar to the conventional mean shift methods, the location of the tumor in the first frame needs to be decided by the user. Mean shift algorithm with multiple features allows estimating the correct center of the tumor in each US frame, which may have been slightly deformed through respiration. Afterwards, we apply active contour version of dynamic directional gradient vector flow (DDGVF) to find the exact boundary of tumors (Jierong et al., 2006). DDGVF is very suitable for low quality intensity based US images, since it has stronger force for absorbing desirable boundaries.

Actually mean-shift allows us to estimate the movement of the target between to successive frames (Comaniciu et al., 2002). By sampling a few points on the boundary of final active contour in the current frame then translating them based on the relocation of the center of ellipse in two consecutive frames, we define the initial points of active contour for the next frame. Figure 1 shows the sketch of our algorithms for the first and second frame (Tashk & Faez, 2007) (Asgari, Shafran, & Bayestehtashk, 2012).
The remainder of this paper is organized as follows. In section 2 the investigated database and the details of the proposed method for tumor tracking is explained in details. Here we consider multiple features extracted from US images, and explore how to feed these features to the mean shift algorithm. Furthermore, DDGVF algorithm is briefly explained. Section 3 covers experimental results of the proposed method. We discuss the results and achievements in section 4. Conclusions and our future work are presented in section 5.

2. Methodology

Our proposed method involves two main steps. In the first step, we apply the mean shift algorithm to the first frame of the video sequence. We use various features of intensity, edge and texture to find the correct ellipsoid from each frame, and the final ellipsoid of mean shift algorithm will be the weighted version of all three ellipsoids computed on each feature vector. In the second step, we apply dynamic direction gradient vector flow version of the active contour to find the target boundary. We exploit the translated result of final active contour as the initial guess of the active contour in the next frame, and minimize its energy in each single grayscale image. As a result, the exact boundary of the target can be found. We estimated the ellipse surrounding the target in each single image by using the target
model and the mean shift algorithm. Since Dynamic Directional Gradient Vector Flow (DDGVF) model of the active contour can effectively select boundaries and distinguish positive from negative boundaries within complicated scenes, we consider this model as an efficient approach for finding boundaries inside pseudo gray images. The proposed algorithm is resistant to the change in aspect ratio. Figure 2 shows the block diagram of the proposed method.

Fig. 2. Block Diagram of the proposed method

2.1. Dataset description

In order to evaluate the proposed method, 2-D US images of the livers of five subjects suffering from liver tumors were considered (De Luca et al., 2013, Petrusca et al., 2013, Preiswerk et al., 2014, Royer et al., 2014). The US images were captured at 20 frames per second, while the patients were being monitored during the evaluation for ablation treatments. The recording length for each subject is around 2 minutes, and contains approximately 2400 frames of 512*512 pixels each. The US images were obtained using Digital Phased Array System (DiPhAs) from Fraunhofer IBMT©. On average, the recordings include one or two respiratory cycles. For each subject from this dataset, the centers of liver tumors were annotated by a clinical specialist in the first frame of US images.

2.2. Motion estimation

2.2.1. Multiple features

In this section we explain how we model the separate features of the elliptical region. All of these models are based on histograms. Since in our US images the appearance of the studied object changes across different frames, histograms can be employed as a helpful tool, as they possess the interesting property which permits some changes in the object appearance without a significant change in the histogram.

2.2.1.1. Intensity features

Since intensity histograms are robust to partial occlusion, and are scale and rotation invariant, the resulting algorithm can efficiently and successfully handle non-rigid deformation of the target as well as the rapidly changing dynamics in complex unknown background. As the first stage, we model the target using a weighted histogram that considers the intensity of the US image and location of the
target. In the first frame we extract the tumor manually using an ellipse defined by its center and axes. As we move off-center, the interior pixels will become less important, thus to compensate for this issue we allocate some weight to the pixels based on their distance from the ellipse center. Since the peripheral pixels are the least reliable (being affected by occlusions and interference) within background, the use of weights will increase the robustness of density estimation in mean shift algorithm (Babaeian et al., 2008, Babaeian et al., 2009b). We use an 8-bin weighted intensity histogram.

### 2.2.1.2. Edge features

During the recording of our US image sequence, the subject was breathing and therefore the intensity of our video sequence changes throughout the successive frames, which occludes the anatomical structures. Consequently, the intensity histogram alone cannot be used as a reliable measure of estimating tumor location. In this section we take into account the edge features as a useful feature for modeling the structure of the tumor in video sequences. We build an edge histogram by estimating the direction of the edge in each frame. Considering a pixel with density \( I \) inside the ellipse, the edge pseudo image is constructed by taking the gradient of the pixels in \( x \) and \( y \) directions. The edge direction \( \theta \) is thus approximated by (1).

\[
\theta_{uv} = \tan^{-1}\left( \frac{\partial I / \partial y}{\partial I / \partial x} \right)
\]  

(1)

We just consider the edges above a predefined threshold. Therefore the weighted edge histogram for target model is obtained by (2).

\[
\hat{q}_{uv} = C \sum_{k=1}^{m} k \left( \left\| x_{k}^{*} \right\|^{2} \right) \delta \left[ b_{v} \left( x_{i}^{*} \right) - u_{v} \right]
\]

(2)

The function \( b_{v}: R^{2} \rightarrow \{1,...,m_{v}\} \) maps the pixel at location \( x_{i}^{*} \) to the index \( b_{v} \left( x_{i}^{*} \right) \) of its bin in the quantized edge feature space. We use an 8-bin weighted edge histogram. The \( C_{v} \) constant is defined in order to satisfy \( \sum_{u_{v}=1}^{m_{v}} \hat{q}_{uv} = 1 \).

### 2.2.1.3. Texture features

Texture, which is an effective feature in many applications, can serve as an observational model for its better perceptional ability compared to factual measurement (Ozyildiz et al., 2002). Although there is no available unique definition for texture, but in most situations it is assumed as a criterion for describing the spatial pixel-level distribution/information, whether it be stochastic or periodic (Manjunath et al., 1996). Since respiration is a periodic process, speckles appear as texture patterns in the US image sequences. These texture patterns demonstrate a high level of correlation across
successive frames. Therefore, the use of texture histogram should reveal important information about the tissue properties.

Recently, multiresolution and multichannel analysis became one of the major tools in texture extraction. This method extracts information about the image in time-frequency domain, and thus provides a powerful tool for the description of similarities among the textures (Virmani et al., 2013). The feature extraction method adopted here is the spatial-frequency method of wavelets built around discrete wavelet transform (DWT). DWT is one of the most useful techniques for multi-resolution image analysis. To represent the texture inside the ellipse, we construct a feature vector as follows. First, a rectangular region was chosen in the center of elliptic with the length and width equal to the major and minor elliptic diameters, respectively, and then the three-level DWT decomposition is applied to generate ten channels from the intensity images. The texture for the \( m^{th} \) channel is computed as (3),

\[
t_m = 1/LW \sum_{i=1}^{L} \sum_{j=1}^{W} |\sigma(i, j)|
\]

where \( L \) and \( W \) are the channel dimensions, and \( \sigma(i, j) \) is the wavelet coefficient at location \((i, j)\). A ten-element feature vector represents texture of the region, and is given by \( \hat{q}_{m} = \{t_1', ..., t_{10}' \} \). Finally, we will have a 10-element feature vector (histogram) as \( \hat{p}_{n}(y) \) of the texture feature for the target candidate is given by:

\[
\hat{p}_{n}(y) = C_t \{t_1'(y), ..., t_{10}'(y)\}
\]

2.2.2. Mean shift

We search an ellipse in a new region within a new frame, which should be most similar to the ellipse encircling the tumor within the previous frame. Supposing that the tumor movement is negligible among two successive frames, it is most likely that the new ellipse will be about the same location of the ellipse in the previous frame. This kind of exhaustive search is time consuming though, and thus it is recommended to use an adequate recursive algorithm. We define a similarity criterion for comparing two Histograms, and to this end we use Bhattacharya distance separately for each feature. By applying the mean shift algorithm, the current location of the object is obtained as (5).

\[
\hat{y}_i = \frac{\sum_{i=1}^{n} x_i w_i g \left( \|y - x_i\|^2 \right)}{\sum_{i=1}^{n} w_i g \left( \|y - x_i\|^2 \right)}
\]
where \( x_i \) is the normalized \( i \)-th pixel location inside the current frame, and \( w_i \) is obtained by (6).

$$w_i = \sum_{\gamma_c} \frac{\dot{q}_i}{\dot{p}_c(\gamma)} \delta \left[ b(x_i) - u \right]$$  \hspace{1cm} (6)

This method takes into account the Bhattacharyyya distance for exerting some weight. Furthermore, by using the smallest distance measure \( \hat{p}(y) \) for each feature, the weight of that feature \( f \) is determined by:

$$\hat{e}_f = \frac{1}{\hat{p}_{\text{min}}} \ , f = 1, ..., F \hspace{1cm} (7)$$

Weights are then normalized such that \( \sum_{f=1}^{F} \hat{e}_f = 1 \) by:

$$\hat{e}_f = \frac{\hat{e}_f}{\sum_{f=1}^{F} \hat{e}_f} \hspace{1cm} (8)$$

Hence, the new tumor location in each iteration within the current frame is calculated independently using (5) for each feature space. The final tumor location in that iteration is therefore given by (9).

$$\hat{y}_{1\text{multiple}} = \hat{e}_c \hat{y}_{1c} + \hat{e}_c \hat{y}_{1e} + \hat{e}_s \hat{y}_{1t}$$  \hspace{1cm} (9)

### 2.2.3. Boundary estimation using active contour

After finding of the correct ellipse in each frame we use a version of active contour (Kass et al., 1988) to find the exact boundary of the tumor region within the video sequences. To this goal, we use the DDGVF snake introduced by (Jierong et al., 2006). With conventional snake, external force is defined as a function of \( I \nabla I \) (The absolute magnitude of image gradient), which is a conventional edge detector. Since this operator lacks the information on direction of gradient, it is incapable of making distinction between positive and negative edges. In order to solve this issue, a new snake model called DDGVF, is proposed to distinguish between the positive and negative edges (Jierong et al., 2006). DDGVF snake is the generalized model of GVF snake, originally proposed by (Chenyang et al., 1998). A boundary is defined as positive if positive step edges exist along the its outward normal (note that the intensity gradient is estimated along the boundary, and is inward directed), similarly a boundary is defined as negative, if negative step edges exist along the outward normal. In the proposed method, a new edge map is used in order to preserve the directional gradient information.
3. Experimental Results

To evaluate the proposed motion tracking algorithm, we applied the method on the US image sequences of five subjects. For each subject, the centers of tumors in the first frame have been initially marked by an expert. We exploit from this information to define an ellipse required for the mean shift algorithm. Applying the mean shift algorithm with multiple features, we are able to find the correct ellipse surrounding the tumor inside the next frame. Here, the mean shift parameters were selected as \( \lambda = 1, \epsilon_0 = 0.05, \text{kernel} = \text{linear}, \alpha = 1 \). More details on the selection of these parameters are given in (Comaniciu et al., 2003, Babaeian et al., 2009a). To find the correct tumor boundary within each US frame, the active contour model is then applied to this ellipse with parameters set to \( \beta = 0, \gamma = 0.6, \sigma = 1, \mu = 0.2 \).

We evaluated the performance of the proposed method both qualitatively and quantifiably. For the quantitative evaluation, we define a tracking error that implicitly correlates to the performance of this method. In the context of tumor tracking, one may encounter two kinds of errors; (1) target pixels falling outside the active contour, (2) the pixels located inside the active contour not belonging to the target. Both types of errors are incorporated in the overall error value by simple aggregation. The utilized contour tracking error percentage is thus estimated as:

\[
\text{Tracking Error\%} = \frac{\text{Aggregated number of pixels generating both types of errors}}{\text{Aggregated number of pixels from target and the contour}} \times 100
\] (10)

Since there are approximately 2400 frames per subject, we randomly select a sample of 10 frames and determine the correct tumor boundary within those frames and compute the tracking error. If a contour exactly matches with its relevant target boundary the error is zero, whereas the maximum error will occur when the contour and its corresponding target share no common pixels. The contour tracking error percentages for 10 selected frames is depicted in figure 3.

![Fig. 3. Tracking Error Percentage in 10 different frames of ultrasound images. At the beginning the error increase but after 30 frames it fluctuates moderately between 9% and 14%.](image-url)
Figure 4 shows the output of the proposed algorithm on 6 different frames of liver US image sequences captured from subject 2, which could provide a qualitative assessment of the proposed algorithm. As seen from the figures the method could find the target boundaries with good accuracy.
Fig. 4. Output of the proposed liver tumor tracking algorithm for sample frames of 1, 20, 40, 60, 80, 100 from subject 2. This subject had three tumors, and the proposed method was applied individually for each of tumors.
4. Discussion

Tumor motion tracking algorithm involves a number of tunable parameters, affecting the output results, and which can be adjusted to achieve optimal performance. Since the proposed algorithm uses multiple features, the experimental results obtained from different US images related to different subjects demonstrate a robust tracking performance with respect to the size and shape of the tumor as well as to the respiratory cycle, which in turn make it superior to other existing tracking methods. Another important aspect of our algorithm is the ability of simultaneously tracking multiple tumors. In fact, since the independent tracking of each tumor is possible, several tumors at a time can be clinically tracked using the proposed method.

Although the proposed motion-tracking algorithm consists of several steps, yet its general computational cost is low and suitable for real-time implementation. Firstly, we utilize three features of DWT coefficients, edge, and intensity histogram, all being linear and very cost effective. Secondly, the mean shift algorithm is applied on these extracted measures. An optimized implementation of the mean shift algorithm tested on a 3GHz PC on a basic framework with scale adaptation, involving three optimizations in each iteration, runs as fast as of 20 fps allowing simultaneous tracking of up to three targets in real time. It should be noted that without scale adaptations these numbers should be multiplied by three, resulting in an even higher frame rate. Furthermore, computational complexity of DDGVF is of the order O(N^3) where N is the number of pixels inside the ellipse found by mean shift algorithm, which for our video sequences approximately varies between 400 to 2500.

Here we applied the method on the US images sequences with 20 fps frame rate. However, as the respiration is not a very fast process, for the sake of clinical application an alternative methodology would be to subsample the captured frames and apply the method to a subset of video sequences, and k m to find the center of tumor in the remaining frames through an interpolation. Furthermore, Using the translated active contour helps us to find the boundary of the object in the next frame with lower number of iterations, which greatly reduce the computational costs.

5. Conclusion

We presented a novel method for real-time tumor motion tracking, by combining mean shift and active contour models to analyze US images. Considering that US images are of lower quality compared to the other types of images, we first extracted multiple features and computed histogram of that feature, and then applied mean shift algorithm to the features of each frame to estimate the location of the tumor in that frame. Afterwards, we applied a DDGVF model of active contour to the output of the mean shift algorithm on each frame to extract the tumor boundary. Since respiration is a periodic phenomenon, a future investigation may consider finding its period and to use this information for achieving a more reliable performance for the proposed algorithm. The accurate and relatively fast performance of our method makes it an ideal solution for real-time clinical applications.
References

Babaeian A, Tashk AB, Barzin F, Hosseini SM. Target Tracking Using Mean Shift and Dynamic Directional Gradient Vector Flow. System Theory, 2008 SSST 2008 40th Southeastern Symposium on2008. p. 366-70.

Babaeian A, Rastegar S, Bandarabadi M, Erza M. Modify kernel tracking using an efficient color model and active contour. System Theory, 2009 SSST 2009 41st Southeastern Symposium on2009a. p. 59-63.

Babaeian A, Rastegar S, Bandarabadi M, Rezaei M. Mean shift-based object tracking with multiple features. System Theory, 2009 SSST 2009 41st Southeastern Symposium on2009b. p. 68-72.

Buzurovic I, Podder TK, Ke H, Yan Y. Tumor Motion Prediction and Tracking in Adaptive Radiotherapy. BioInformatics and BioEngineering (BIBE), 2010 IEEE International Conference on2010. p. 273-8.

Chenyang X, Prince JL. Snakes, shapes, and gradient vector flow. Image Processing, IEEE Transactions on. 1998;7:359-69.

Cifor A, Risser L, Chung D, Anderson EM, Schnabel JA. Hybrid Feature-Based Diffeomorphic Registration for Tumor Tracking in 2-D Liver Ultrasound Images. Medical Imaging, IEEE Transactions on. 2013;32:1647-56.

Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2002;24:603-19.

Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2003;25:564-77.

De Luca V, Székely G. Liver Motion Tracking in Ultrasound Sequences for Tumor Therapy: ETH-Zürich; 2013.

Goyal S, Kataria T. Image Guidance in Radiation Therapy: Techniques and Applications. Radiology Research and Practice. 2014;2014:10.

Jierong C, Foo SW. Dynamic directional gradient vector flow for snakes. Image Processing, IEEE Transactions on. 2006;15:1563-71.

Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. Int J Comput Vision. 1988;1:321-31.

Manjunath BS, Ma WY. Texture features for browsing and retrieval of image data. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1996;18:837-42.

Ozyilidi E, Krahnstöver N, Sharma R. Adaptive texture and color segmentation for tracking moving objects. Pattern Recognition. 2002;35:2513-29.

Petrusca L, Cattin P, De Luca V, Preiswerk F, Celicanin Z, Auboioux V, et al. Hybrid ultrasound/magnetic resonance simultaneous acquisition and image fusion for motion monitoring in the upper abdomen. Investigative radiology. 2013:48:333-40.

Preiswerk F, Arnold P, Fasel B, Cattin PC. Robust tumour tracking from 2D imaging using a population-based statistical motion model. Mathematical Methods in Biomedical Image Analysis (MMBIA), 2012 IEEE Workshop on2012. p. 209-14.

Preiswerk F, De Luca V, Arnold P, Celicanin Z, Petrusca L, Tanner C, et al. Model-guided respiratory organ motion prediction of the liver from 2D ultrasound. Medical Image Analysis. 2014;18:740-51.

Royer L, Babel M, Krupa A. Non-rigid target tracking in 2D ultrasound images using hierarchical grid interpolation. 2014. p. 90343O-O-7.

Rubin JM, Feng M, Hadley SW, Fowlkes JB, Hamilton JD. Potential use of ultrasound speckle tracking for motion management during radiotherapy: preliminary report. Journal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine. 2012;31:469-81.

Sharp GC, Jiang SB, Shimizu S, Shirato H. Prediction of respiratory tumour motion for real-time image-guided radiotherapy. Physics in Medicine and Biology. 2004;49:425.

Shimizu S, Shirato H, Ogura S, Akita-Dosaka H, Kitamura K, Nishioka T, et al. Detection of lung tumor movement in real-time tumor-tracking radiotherapy. International Journal of Radiation Oncology*Biology*Physics. 2001;51:304-10.

Shirato H, Shimizu S, Kunieda T, Kitamura K, van Herk M, Kagei K, et al. Physical aspects of a real-time tumor-tracking system for gated radiotherapy. International Journal of Radiation Oncology*Biology*Physics. 2000;48:1187-95.

Virmani J, Kumar V, Kalra N, Khandelwal N. SVM-based characterization of liver ultrasound images using wavelet packet texture descriptors. Journal of digital imaging. 2013;26:530-43.

Youngkyoo H, Jung-Bae K, Won-Chul B, Kim JDK, Chang-Yeong K, Heesae L. Robust real-time respiratory motion tracking using ultrasound image sequences. Ultrasونics Symposium (IUS), 2012 IEEE International2012. p. 1666-9.

Asgari, M.; Shafran, I.; Bayestehtashk, A.; "Robust detection of voiced segments in samples of everyday conversations using unsupervised HMMS," in Spoken Language Technology Workshop (SLT), 2012 IEEE , vol., no., pp.438-442, 2-5 Dec. 2012

Tashk, A.R.B.; Faez, K., "Boosted Bayesian Kernel Classifier Method for Face Detection," in Natural Computation, 2007. ICNC 2007. Third International Conference on , vol.1, no., pp.533-537, 24-27 Aug. 2007