Registration of ultrasonography sequences based on temporal regions

Sebastian Schäfer*, Paolo Angelelli†, Kim Nylund‡§, Odd Helge Gilja‡§ and Klaus Tönnies*
*Department of Simulation and Graphics, University of Magdeburg, Germany,
Email: sebastian.schaefer@ovgu.de
†Department of Informatics, University of Bergen, Norway
‡Institute of Medicine, University of Bergen, Norway
§National Centre of Ultrasound in Gastroenterology, Haukeland University Hospital, Bergen, Norway

***Abstract*—2D Ultrasonography images with parallel contrast enhanced sequences for analysis constitute a rapid and inexpensive imaging technique with high temporal resolution to assess perfusion of tissue. However, motion from various influences corrupts the inter-pixel correspondences between different time frames and therefore hampers computer-assisted analysis of perfusion parameters. We present a user-supported method applying a temporal similarity matrix to remove frames with uncorrectable out-of-plane motion. For the remaining regions of frames, motion influence can be compensated for by image registration. Subsequently B-Spline based registration is applied using the temporal regions with automatic determination of a suitable reference frame image. Evaluation with ground truth data of six datasets comparing a medical expert frame analysis to the proposed technique yields 85.1 % sensitivity and 81.7 % specificity in average. On average 6 % of the frames have been erroneously included in temporal regions, although they contain out-of-plane motion.**

I. Introduction

2D ultrasonography (US) is one of the most widespread medical imaging technique. It enables immediate and inexpensive examinations with high spatial and temporal resolution and no contraindications or radiation exposure. US is also used for perfusion imaging employing contrast agents, consisting of gas-filled microbubbles that have a high degree of echogenicity [1], [2]. This has become an excellent tool for delineating the vascular structure in normal and pathological tissue, in order to detect primary tumours and metastases [3] in various organs, like liver [4] and pancreas [5]. The perfusion analysis is performed by extracting and understanding the perfusion kinetics of the blood in the tissue of interest from the acquired multi-frame data. This is possible after the injection of the contrast agent, that makes the blood traceable [6].

During 2D contrast enhanced US (CEUS) examinations studying perfusion, the sonographer normally will hold the probe still in a particular position and orientation, to image a suitable slice of tissue of interest during contrast agent administration. However, the data acquired with this examination methodology often contains significant motion. The reason is that patient movement through breathing and differently induced movements (intrinsic induced motion) are present in addition to the motion caused by tilting or moving the US probe (extrinsic induced motion), since US imaging is normally performed handheld.

While this motion can normally be interpreted by well trained physicians [7], in computer-assisted analysis the different image frames of a time-dependent acquisition are required to be kept aligned in order to extract the signal intensity of a voxel at different time points. Aligning this stack of time-dependent cross sections is complicated by the fact that the 2D US image plane can "miss" the region of interest during part of the examination, due to the three dimensional nature of the motion described above. Therefore, we distinguish between two types of motion effects on the images:

- *correctable image movement*, when the image plane contains the region of interest (ROI), but this region is moved or deformed with respect to a reference image
- *uncorrectable image movement*, when the US image plane does not contain the ROI at all

In performing computer-assisted analysis of CEUS data, frames not imaging the ROI should be excluded from analysis, as they would distort the results, while the others should be aligned in order to improve the pixel correspondences between different time steps [8].

In this paper, we present a two-stage approach for registration of 2D US sequences. The first stage is designed to assist the user to identify those frames containing uncorrectable motion from the acquired image stack and group temporal regions with frames where differences can be described with in-plane motion only. In the second stage, automatic registration of the remaining data is performed. This stage produces a registered sequence within the bounds of the identified temporal regions. This has to be taken into account in later analysis (Fig 1).

II. Related Work

Registration of mono-modal image sequences is a common task in medical image analysis [9]. For motion analysis, e.g. to assess elastic and contractile properties of the myocardium [10], only consecutive time frames are registered. In most other cases, sequence images are registered to a predefined reference frame and registration is performed including all available time frames. Kolár et al. [11] investigate the ability of mutual information as similarity measure to cope with the contrast change of objects due to contrast agent administration.

When 2D data is used, images acquired at different times may contain mixed information from different planes. Conse-

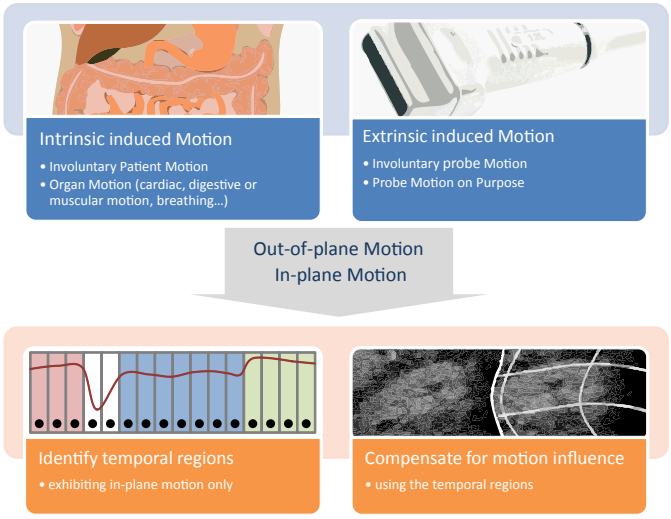


Fig. 1: Different types of motion influence in CEUS image sequences

quently, registration of each frame to the reference image to compensate for motion is only advisable if the same image plane is depicted and thus, registration is able to produce a valid result at all.

There are few approaches dealing with detection of possible outliers or instances representing different content within an image sequence. Renault et al. [7] propose a frame selection method to identify frames belonging to the same point in time of the respiratory cycle by using independent component analysis to separate the influence of motion from the influence of contrast uptake. An essential assumption is that no motion other than breathing motion is present, because objects in the scene must reach almost the same position at each time point in the cycle. Frouin et al. [12] compensate for motion influence in dynamic arterial perfusion studies. Registration of image pairs is rejected if the standard deviation in a predefined region of the difference image exceeds a certain threshold to prevent registration of images with large motion shifts.

In contrast, we propose the identification of regions of frames representing the scene with similar probe orientation and containing, at the most, in-plane or cyclic occurring motion. These could be consecutive time frames or time frames from different positions during acquisition.

III. Method

In CEUS imaging, datasets are acquired using brightness modulation (b-mode) and contrast modulation[1] alternately (Fig. 3a and 3b). Thus each frame has a b-mode and contrast-mode representation acquired at the same time. Both image sequences usually have a temporal resolution around 10 frames per second. In this section the identification of temporal regions and the registration for motion compensation are described. The motion analysis and calculation of registration transformations are computed using the b-mode data containing anatomical information. Results can then be applied to the contrast sequence that represent the perfusion information.

[1]modality targeted at the detection of US contrast agent

A. Identification of related frames

The goal of the identification step is to detect subsequences of frames acquired without changing the probe position exhibiting in-plane motion only. The identification of those temporal frames is based on normalized correlation (NC, eq. 1) that determines the similarity of time frames A and B with i being the pixel index and N the amount of total pixels.

$$NC(A,B) = \frac{\sum_{i=1}^{N}(A_i \cdot B_i)}{\sqrt{\sum_{i=1}^{N} A_i^2 \cdot \sum_{i=1}^{N} B_i^2}} \quad (1)$$

This measure has been widely used for similarity determination for image registration [9] using image intensity itself or extracted features. It is also applied in case of US image registration [7], [13] as it is robust against noise and allows for relative comparison of values between time frames, as it produces values in the interval [1, -1]. In most cases, analysis is directed towards a particular region of interest (organ or tissue region). Therefore, a mask for calculating similarity of frames can be defined manually.

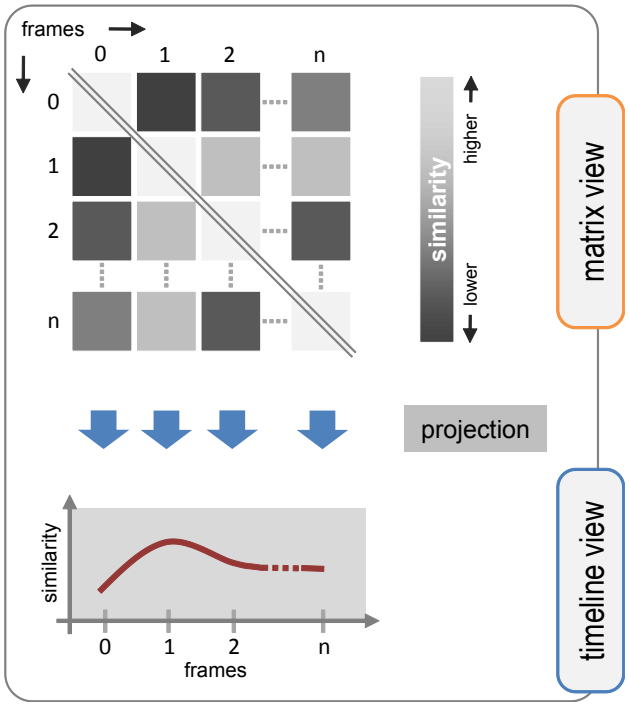


Fig. 2: The matrix view depicts inter-frame similarity of all available frames. Brighter color means higher similarity. The timeline view is a projection (average) of the matrix columns (or rows) representing the average similarity of each frame to all other frames.

For the analysis of frame similarity two representations have been developed. For the matrix view (Fig. 2) the similarities of each time frame to all other frames of the dataset are computed

and depicted in a co-occurrence matrix plot (examples see Fig. 3c-3e). Irregularly occurring square-shaped characteristics (Fig. 3d, 3e) are indicative for extrinsic motion influences, in particular caused by motion of the US probe and thus the projection direction. Regularly occurring patterns in the matrix which are visible in Fig. 3c and less strongly pronounced in Fig. 3d are stemming from cyclic motion mostly caused by heart beat. Often this represents the motion which should be compensated, to establish valid perfusion measurement in a specific organ or tissue type. In contrast, Fig. 3e does not exhibit a regular occurring pattern as the related image depicting a kidney is not influenced by motion induced by heart beat.

To generate temporal regions of related frames, a projection (timeline view, Fig. 2) is derived by averaging the matrix row/column data into a 1D curve (Fig. 3i, 3j).

A watershed-based segmentation with a user set parameter δ defining equidistant scanlines is used to divide segments of the curve into different corridors (Fig. 3i) to automatically generate the temporal regions (in ref. to Tab. II). Parts of the curve lying in the same corridor are assigned to the same region although they may not be connected temporally. Regions with less than 10 % of the total number of frames are omitted, as they will not contribute to functional diagnosis. By reducing the scanline distance δ interactively, the intra regional similarity and thus the number of required temporal regions increases.

Furthermore, generated regions are transferred to the similarity matrix plot. Subsequently, the user can edit the proposed result on the basis of visual exploration of the similarity co-occurrence plot (Fig. 3g) which represents a more detailed view on the frame relations. It is also possible to manually define a new set-up of regions or to interactively separate and merge existing ones. Regions can be adapted easily, using the information provided by the matrix and timeline view and the automatic region extraction. We refer to this approach as semi-automatic (Tab. II).

Once the regions are defined, a reference image for each region is determined by using the frame with highest average similarity to all respective other frames in the same region in order to minimize the size of required transformations. This frame is used as reference image in the subsequent registration process.

B. Registration

In the second stage of our pipeline from Fig. 1, the registration is performed considering the aforementioned regions of frames and their respective reference images.

In all registration steps, we use the NC measure as quality of fit of determined transformations between all frames of a region and the computed regional reference image. As in the previous step the same ROI mask can be used to perform registration and calculate the similarity of images.

In-plane motion to be corrected within identified temporal regions may be composed of an arbitrary combination of linear shifts due to patient or probe motion. Cyclic occurring motion causes non-linear deformations due to intrinsic motion. To compensate for linear shifts, a rigid registration step allowing translations and a rotation around the centre of the image can be performed to pre-align the frames. This avoids unnecessary distortion of the subsequent non-rigid registration using a B-Spline-based transformation to compensate for non-linear types of deformations.

We use a 5×5 point grid to represent cubic B-Spline functions on the image. Additional points are placed on the outside of the image to support calculation of the inner point region resulting in an 8×8 point grid. A bounded limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm is used to optimize the 64 point locations (128 parameters), whereas we constrain each point to only move within half of the spacing of the grid points, to disallow degeneration of the grid. The transformation parameters of each frame registration are initialized with the final transformation parameters from the preceding frame for reasons of stability and efficiency.

IV. Results and Discussion

For evaluation, six datasets showing the right iliac artery, neo-terminal ileum, liver, pancreas or kidney have been used with a spatial resolution ranging between 200 and 400 pixels in x- and y-direction and a temporal resolution between 200 and 700 frames. To ensure that our approach works properly, we expect datasets consisting of 40 % frames with in-plane motion only and a maximum of 5 different temporal groups of frames, in order to be differentiated with the co-occurrence similarity representation.

As a proof of concept, we measured the average standard deviation (STD) of all pixels over the different time steps in the datasets a) unregistered, b) after applying our region-based registration and c) after a registration covering all frames of the dataset and registered to one fixed frame image (Tab. I). STD is used as control measurement, as it expresses the size of variation when the time frame images are composed of an equal level of intensity. The average STD was measured in the defined regions used for the region-based registration only, to compare the performance of both strategies.

TABLE I: Evaluation of the standard deviaton within the regions of interest before registration, after registration of temporal frames to the reference frame of the region and after registration of all frames to one still frame. Temporal regions have been generated with the automatic method.

	before	temporal reg. (%[2])	all frames reg. (%)
dataset 1	7.13	5.61 (27 %)	5.67 (26 %)
dataset 2	11.84	7.29 (62 %)	7.85 (51 %)
dataset 3	15.66	11.98 (31 %)	12.95 (21 %)
dataset 4	13.22	9.45 (40 %)	9.77 (35 %)
dataset 5	18.61	12.70 (47 %)	15.82 (18 %)
dataset 6	9.16	7.34 (25 %)	7.62 (20 %)
average		**9.06 (38.5 %)**	9.95 (33.0 %)

[2]improvement compared to STD before registration

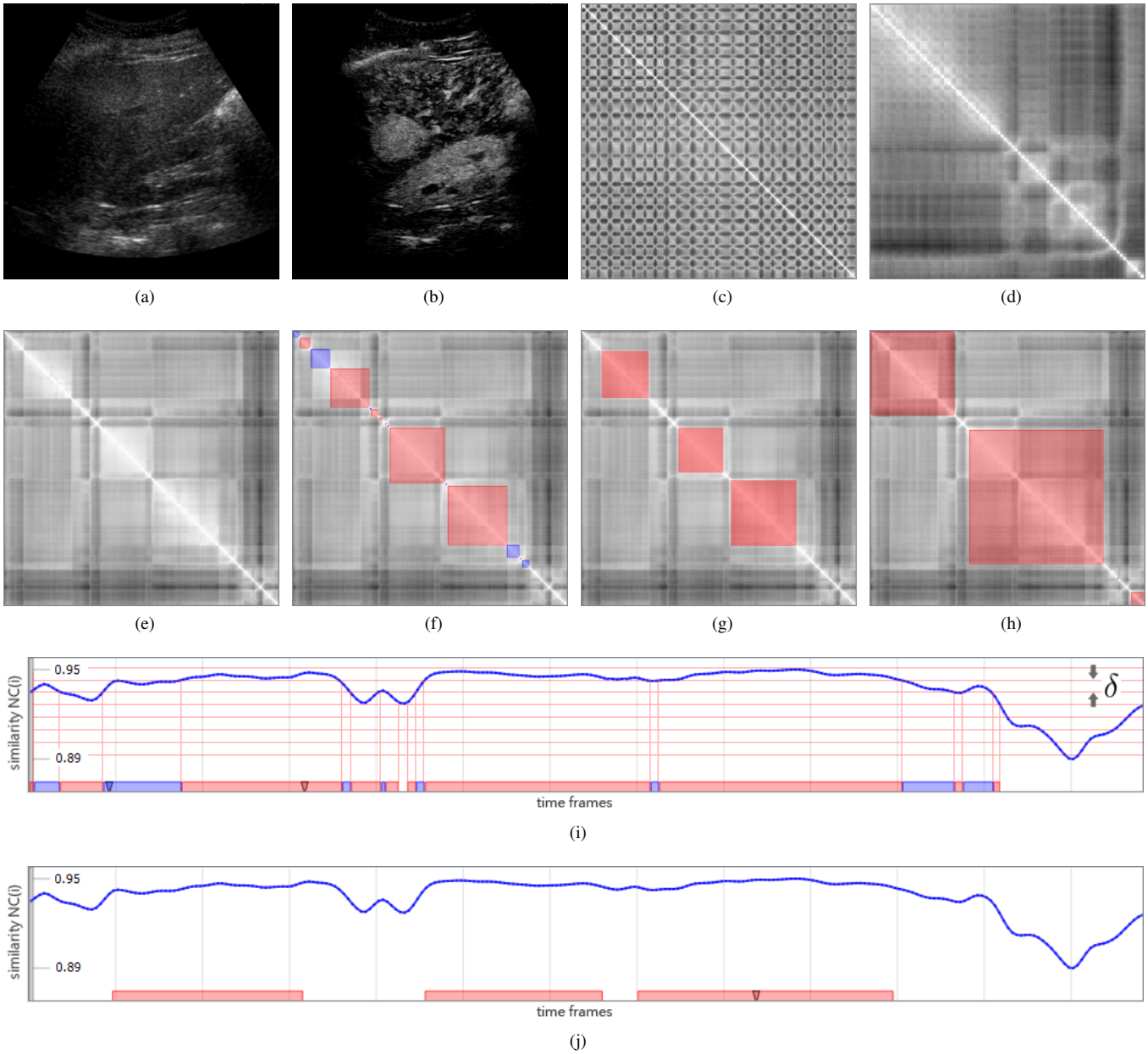
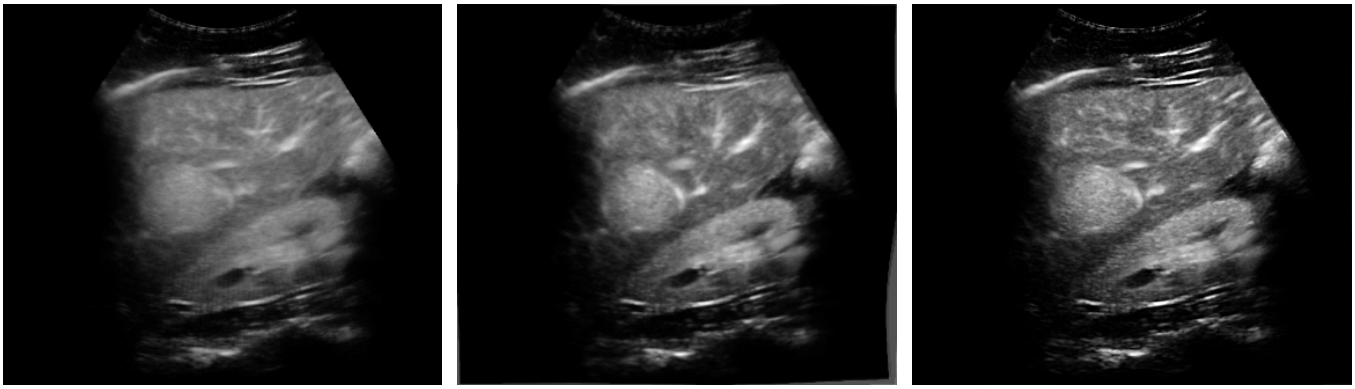


Fig. 3: (a) and (b) are single frames of US images, b-mode and contrast modulation, respectively. (c) and (d) depict similarity co-occurrence matrices of two different datasets. (f) represents automatically generated regions for dataset shown in (a) from the derived timeline plot (i). (g) shows the regions adapted by the user. (h) represents ground truth data for the dataset in (a). (i,j) projections of average similarities with regions from (f) and (g). Triangular markers indicate the reference image of the region to be used.

For region-based registration it exhibits at least equivalent or better values compared to a registration taking all frames into account (avg. 9.8 %). This implies that registration of temporal regions with the specially determined reference image yields higher accuracy in terms of a reduction of intensity variation at pixel locations compared to registration of all frames of the sequence using just one reference image. Additionally, the mean image of a sample dataset is shown before registration (Fig. 4a), after standard registration to a single reference frame (Fig. 4b) and our proposed temporal region-based registration (Fig. 4c). Contours are clearer towards the right, indicating better registration fits.

In a second investigation, we compared ground truth data to the result of the temporal regions defined by the proposed method. The ground truth data was generated through in detail analysis of the datasets by medical experts. They identified out-of-plane frames manually and grouped the remaining frames exhibiting in-plane motion only. The overlap between



(a) (b) (c)

Fig. 4: Mean of Contrast image sequence before registration (a), after a global registration to a single reference frame (b) and after temporal region registration (c). Only registered frames are included in mean image (for b and c).

TABLE II: Evaluation results for the six datasets: temporal regions compared to ground truth data. Overlap in percent between both and the amount of frames included in the analysis although they exhibit out-of-plane motion (overseg) and the frames which have been left out (underseg) by our method.

	overseg (%)	underseg (%)	overlap (%)
dataset 1			
automatic	12.8%	5.5%	94.5%
semi-automatic	0.9%	30.6%	69.4%
dataset 2			
automatic	0.0%	11.4%	88.6%
semi-automatic	0.0%	11.6%	88.4%
dataset 3			
automatic	19.9%	1.8%	98.2%
semi-automatic	7.7%	6.1%	93.9%
dataset 4			
automatic	25.9%	19.0%	81.0%
semi-automatic	25.9%	19.0%	81.0%
dataset 5			
automatic	0.0%	15.2%	84.8%
semi-automatic	0.0%	20.2%	79.8%
dataset 6			
automatic	1.9%	3.8%	96.2%
semi-automatic	0.0%	2.9%	97.1%
average			
automatic	10.1%	9.4%	90.6%
semi-automatic	5.8%	15.1%	84.9%

automatically generated regions and the ground truth data is calculated counting all frames, which have been declared to belong to a temporal region in both.

Evaluation yields 91 % average overlap for automatic frame selection with 10 % of the frames being erroneously declared to belong to an in-plane motion subsequence. This is mainly caused by datasets 3 and 4 (Tab. II) with a high level of different cyclic occurring motion yielding more than 20 different temporal regions and thus, does not meet our initial assumptions. For the semi-automatic approach an average overlap of 85 % is achieved. The automatic approach generates more regions with a size over 10 % of the total number of frames. Thus the coverage is higher compared to the semi-automatic method. However, the amount of erroneously segmented frames is lower (6 %) using the semi-automatic method.

On average, 9 % of the frames (automatic) and 15 % (semi-automatic) have not been included in an in-plane motion subsequence by our method, although they have been included in the ground truth data. Thus, the temporal regions to be registered and analysed are simply smaller, but are lying within ground truth temporal regions (Fig. 3g and 3h) and most notably do not contain data of other planes. As the semi-automatic approach includes manually defined regions and users tend to keep a safety gap, this number is higher for the semi-automatic method.

V. Conclusion

We proposed a semi-automatic method for US sequences to remove frames with uncorrectable motion and identify regions of frames where motion influence can be corrected by registration. The automatic calculation of frames is proposed to the user, who can adjust, merge and change the regions by the help of a co-occurrence similarity matrix. This helps to establish valid temporal alignment between time frames to allow for quantification and diagnostic purposes. In comparison to a detailed frame analysis, our method is much faster.

Good accordance was achieved when comparing the results to ground truth data (acquired by an in detail analysis by medical experts). After establishing the temporal regions, registration is performed within these regions using the frame with highest similarity to all other region frames as a reference image. This ensures the compensation for motion artefacts in areas where a valid inter-pixel correspondence can be established.

The matrix view provides a more detailed overview of frame similarities. Currently, the definition of temporal regions from the matrix view requires interaction with the user, although region are initially proposed by an automatic approach. If there are more than 5 temporal regions, a lot of manual work has to be done to select and group the frames. It is desirable, to extend the automatic method to work on the co-occurrence matrix view.

Moreover, we plan to integrate contrast images in the calculation of registration to improve the quality of fit especially in areas where contrast enhancement is present. Therefore, assumptions about perfusion characteristics could help to guide registration and the determination of transformation quality in particular [14].

ACKNOWLEDGEMENT

This work has been funded by DFG (research grant no. TO166/13-1).

REFERENCES

[1] M. Postema and O. Gilja, "Contrast-enhanced and targeted ultrasound," *World Journal of Gastroenterology*, vol. 17, no. 1, p. 28, 2011.

[2] T.G. Leighton, *The acoustic bubble*. Academic Press, 1997.

[3] D. Klein, M. Jenett, H. Gassel, J. Sandstede, and D. Hahn, "Quantitative dynamic contrast-enhanced sonography of hepatic tumors," *European Radiology*, vol. 14, no. 6, pp. 1082–1091, 2004.

[4] E. Quaia, F. Calliada, M. Bertolotto, S. Rossi, L. Garioni, L. Rosa, and R. Pozzi-Mucelli, "Characterization of focal liver lesions with contrast-specific us modes and a sulfur hexafluoride-filled microbubble contrast agent: diagnostic performance and confidence," *Radiology*, vol. 232, no. 2, p. 420, 2004.

[5] M. Kitano, H. Sakamoto, U. Matsui, Y. Ito, K. Maekawa, T. von Schrenck, and M. Kudo, "A novel perfusion imaging technique of the pancreas: contrast-enhanced harmonic EUS," *Gastrointestinal Endoscopy*, vol. 67, no. 1, pp. 141–150, 2008.

[6] R. Kolár, R. Jirík, V. Harabis, M. Mézl, and M. Bartos, "Advanced methods for perfusion analysis in echocardiography." *Physiological research/Academia Scientiarum Bohemoslovaca*, vol. 59, p. 33, 2010.

[7] G. Renault, F. Tranquart, V. Perlbarg, A. Bleuzen, A. Herment, and F. Frouin, "A posteriori respiratory gating in contrast ultrasound for assessment of hepatic perfusion." *Physics in Medicine and Biology*, vol. 50, no. 19, pp. 4465–80, 2005.

[8] P. Angelelli, K. Nylund, O. Gilja, and H. Hauser, "Interactive visual analysis of contrast-enhanced ultrasound data based on small neighborhood statistics," *Computers & Graphics*, vol. 35, no. 2, pp. 218–226, 2010.

[9] J. Maintz and M. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[10] M. Ledesma-Carbayo, J. Kybic, M. Desco, A. Santos, and M. Unser, "Cardiac motion analysis from ultrasound sequences using non-rigid registration," in *Medical Image Computing and Computer-assisted Intervention (MICCAI)*, vol. 32, no. 4, 2010, pp. 889–896.

[11] Kolár, R. and Jirík, R. and Harabis, V. and Nylund, K. and Gilja, O.H., "Registration of ultrasound contrast images for perfusion analysis," in *Ultrasonics Symposium (IUS), 2009 IEEE International*. IEEE, 2009, pp. 1251–1254.

[12] F. Frouin, S. Duteil, D. Lesage, P. Carlier, A. Herment, and A. Leroy-Willig, "An automated image-processing strategy to analyze dynamic arterial spin labeling perfusion studies. Application to human skeletal muscle under stress." *Magnetic Resonance Imaging*, vol. 24, no. 7, pp. 941–51, 2006.

[13] J. Luo and E. Konofagou, "A fast normalized cross-correlation calculation method for motion estimation." *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 57, no. 6, pp. 1347–57, 2010.

[14] S. Schäfer, C. Hentschke, and K. Tönnies, "Local similarity measures for ROI-based registration of DCE-MRI of the breast," in *Proceedings of Medical Image Understanding and Analysis (MIUA)*, Warwick, UK, 2010, pp. 159–163.