This document is a manuscript of:

Piotr M. Szczypiński and Artur Klepaczko.
"Deformable Mesh for Regularization
of Three-Dimensional Image Registration."
International Conference on Information Technologies in Biomedicine.
Springer, Cham, 2019.

*https://doi.org/*10.1007/978 − 3 − 030 − 23762 − 2_7

Source codes:
https://gitlab.com/piotr.szczypinski/deformowalne

ITIB web page:
https://itib.polsl.pl/

Deformable mesh for regularization of three-dimensional image registration

Piotr M. Szczypiński and Artur Klepaczko

Institute of Electronics, Lodz University of Technology
piotr.szczypinski@p.lodz.pl, artur.klepaczko@p.lodz.pl

Abstract. We demonstrate that the registration of three-dimensional medical images demands elastic transformation with limitation of scaling. Moreover, it should be performed on a limited number of feature points and within a specified volume of interest. The regularization term derived from locally estimated unimodal transformation meets these requirements. The term was implemented in the image registration routine in the way inspired by deformable models. The resulting algorithm is presented in detail and verified on time series of three-dimensional kidney images. The qualitative results are presented and compared with the registrations obtained by the reference state-of-the-art method.

Keywords: image registration, deformable mesh

1 Introduction and goal

In ill-posed problems, particularly when solving inverse problems, the unique solution may not exist. We usually deal with a number of good or sub-optimal solutions. Regularization is a way of introducing additional information that enables narrowing the range of solutions or indicate the one which is optimal. The optimal means that the solution fulfills a regularizer (or regularization term), which is usually defined as a supplemental goal function. In biomedical applications such a term may introduce an a-priori knowledge on the anatomy or on mechanisms occurring in living organisms.

Image registration [2] is a process of fitting together a content of two or more images. In medical applications it is required to correctly overlap or integrate data obtained by various imaging modalities. An example is to combine Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) with Positron Emission Tomography (PET), where CT and MRI provide anatomical information and PET indicates location of lesions. Another use of the registration is to follow specific anatomical structures in a series of images acquired in time. The structures may change their location and orientation due to respiratory activity or other unavoidable motions.

First of all, having two images to be registered, it is required to identify pairs of corresponding points – one point located in the first image and the other point in the second one. Pairing or finding corresponding points, involves comparison of their neighborhoods in terms of similarity measure. If the similarity is high

enough the two points are paired. The bad thing is that the localization of such points is usually inaccurate due to image discrete form, inherent noise or repeated image patterns which cause ambiguity in the assignment. Therefore, the image registration is indeed an inverse and ill-posed problem, which requires regularization.

There are three approaches [2] to solve the regularization problem in image registration. The first one assumes that images present the same rigid structure and thus translation of one image with respect to the other is limited to rotation and translation. The second approach assumes that the misalignment between the images can be corrected by an affine transformation. Comparing with the first approach, the affine transformation also enables resizing and shear (directional stretching). The third approach enables local image deformation, referred as curved or elastic image transformation. In contrast to the affine transform, which preserves straight lines and keep them parallel, the curved transform allows for bending. In all the listed approaches, the registration algorithm seeks the transformation parameters to minimize some goal function. An example of a goal function may be the mean squared distance between paired points after the transformation have been applied.

Finally, the extra fourth approach does not introduce any regularization term – it assumes that all the paired points are perfectly identified without any localization error. In this approach, one of the images is warped to exactly align the points with their counterparts in the other image.

The human or animal body deformation due to respiratory process have specific properties. The organs may be shifted, slightly rotated or locally deformed. The deformation may include stretching or bending of structures. However, since most of organs are built of cells containing incompressible water and, in a short time range, the water content in cells can be assumed constant, the deformation is generally volume-preserving.

Having these properties in mind we can conclude that the regularization based on the rigid or the affine transformations should be rejected as not allowing for local deformations. Moreover, the affine transformation does not restrict resizing – it is not volume-preserving. Not using any regularization term is also inadequate. Incorrect or erroneous matching of points without proper regularization may lead to unacceptable and unrealistic image deformations. Therefore, the only acceptable solution is the third approach, the elastic image transform. However, also in this case one can find that the most popular way of deformation or elasticity modeling, involving b-splines or polynomials of a limited degree, does not provide a sufficient control over what aspects of the deformation are enabled and which are restricted.

The goal of this work is to present a method for regularization to be applied in three-dimensional image registration. The method gives control to arbitrarily and independently restrain or enable such aspects of image transformation as resizing, stretching, bending and rotation.

2 Transformation from a set of paired points

The affine transformation (1) in three-dimensional space is represented by a 3×3 matrix (rotation, shear and scaling), and a three-element translation vector. If there are 4 noncoplanar points, which coordinates are known before ($\mathbf{w}$) and after ($\mathbf{v}$) transformation, then the transformation matrix $\mathbf{J}$ and vector $\mathbf{T}$ can be uniquely determined. If there are $P > 4$ of such points (Fig.1), the problem becomes ill-posed and requires regularization to be solved. To do this, we assume that the coordinates measured after the transformation include some error. In this case we seek the solution to minimize the mean squared error ε. The error is a function of the elements of the transformation matrix and vector (2). The partial derivatives (3) and (4) of the error function should be equal to zero to minimize the error. The resulting system of equations has a unique solution (5). More detailed reasoning and derivation of this solution can be find in [5].

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_D \end{bmatrix} \approx \begin{bmatrix} j_{11} & \cdots & j_{1D} \\ & \vdots & \\ j_{D1} & \cdots & j_{DD} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_D \end{bmatrix} + \begin{bmatrix} t_1 \\ \vdots \\ t_D \end{bmatrix} = \mathbf{J}\mathbf{w} + \mathbf{T} \tag{1}$$

$$\varepsilon = \sum_{p=1}^{P} \left(|\mathbf{J}\mathbf{w}_p + \mathbf{T} - \mathbf{v}_p|^2 \right) = \sum_{p=1}^{P} \sum_{n=1}^{D} \left(\sum_{m=1}^{D} (j_{nm} w_{pm}) + t_n - v_{pn} \right)^2 \tag{2}$$

$$\frac{\partial \varepsilon}{\partial t_k} = 2 \sum_{p=1}^{P} \left(t_k + \sum_{m=1}^{D} (j_{km} w_{pm}) - v_{pk} \right) = 0 \tag{3}$$

$$\frac{\partial \varepsilon}{\partial j_{kl}} = 2 \sum_{p=1}^{P} \left(w_{pl} \sum_{m=1}^{D} (w_{pm} j_{km}) + w_{pl} t_k - w_l v_{pk} \right) = 0 \tag{4}$$

$$\begin{bmatrix} j_{k1} \\ \vdots \\ j_{kD} \\ t_k \end{bmatrix} = \begin{bmatrix} \sum_{p=0}^{P} w_{p1}^2 & \cdots & \sum_{p=0}^{P} w_{p1} w_{pD} & \sum_{p=0}^{P} w_{p1} \\ \vdots & & \vdots & \vdots \\ \sum_{p=0}^{P} w_{p1} w_{pD} & \cdots & \sum_{p=0}^{P} w_{pD}^2 & \sum_{p=0}^{P} w_{pD} \\ \sum_{p=0}^{P} w_{p1} & \cdots & \sum_{p=0}^{P} w_{pD} & P \end{bmatrix}^{-1} \begin{bmatrix} \sum_{p=0}^{P} w_{p1} v_{pk} \\ \vdots \\ \sum_{p=0}^{P} w_{pD} v_{pk} \\ \sum_{p=0}^{P} v_{pk} \end{bmatrix} \tag{5}$$

As presented in (5), the problem of estimating the affine transformation from a set of paired points can be solved analytically. It requires calculation of the inverse matrix, which in three-dimensional case ($D = 3$) is trivial.

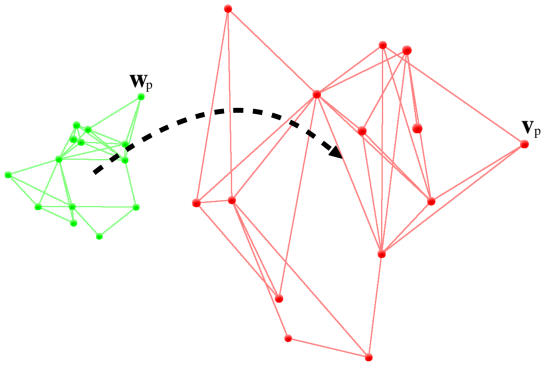

Fig. 1. The goal is to estimate a transformation of green points (LHS) on to the corresponding red points (RHS) with a minimum error.

As aforementioned, the affine transformation defines scaling, shear and rotation of the image space. To establish contribution of these components in the transformation, matrix $\mathbf{J}$ has to be decomposed. The matrix is decomposed to orthogonal $\mathbf{U}$ and symmetric $\mathbf{S}$ matrices (7). Next, the symmetric matrix $\mathbf{S}$ is eigendecomposed to the orthogonal matrix $\mathbf{Q}$ and diagonal matrix $\mathbf{D}$.

$$\mathbf{J} = |\mathbf{J}|^{-\frac{1}{D}}\mathbf{US} \tag{6}$$

$$\mathbf{J} = \mathbf{US} = \mathbf{UQ}^{-1}\mathbf{DQ} = |\mathbf{D}|\mathbf{UQ}^{-1}\begin{bmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_D\end{bmatrix}\mathbf{Q} \tag{7}$$

Determinant $|\mathbf{D}|$ represents scaling, matrix $\mathbf{U}$ defines rotation, and eigenvalues λ_n determine shear at directions indicated by the column vectors of matrix $\mathbf{Q}$. Both $\mathbf{U}$ and $\mathbf{Q}$ matrices are orthogonal, what particularly means that $\mathbf{Q}^{-1} = \mathbf{Q}^{T}$ and $|\mathbf{Q}| = |\mathbf{U}| = 1$.

If all the eigenvalues are equal to one ($\lambda_n = 1$), the transformation does not involve shear and is called a Procrustes transformation. If $|\mathbf{D}| = 1$ then the transformation is volume preserving. However, it must be noted that computation of the affine transformation and then removal of the $\mathbf{U}$ matrix from (7) yields only a rough estimation of Procrustes transformation. The way to exactly estimate the Procrustes transformation is presented in [5] – it is a numerical solution and it is much more demanding computationally.

In contrast to the Procrustes transformation, the volume preserving transformation can be estimated based on the solution (5) by removal of the $|\mathbf{D}|$ factor from equation (7). The above approach to determine the affine transformation and then remove selected factors gives control over the regularizer, which is needed in medical image registration.

3 Image similarity

Finding a correspondence between points in two images requires definition od a similarity function. Selecting a suitable similarity function is not a trivial decision and depends on the properties of the images to be co-registered. The simplest and most straight-forward method to establish similarity between image fragments is the mean absolute difference (MAD). A block of voxels from one image is compared with the block of the same size from the other image. Absolute difference between corresponding pixels in blocks is computed and averaged. The value of this measure is lower if the similarity between blocks is higher. However, this method can be applied exclusively for images of the same modality and acquired under the same conditions, which is a rare case in medical applications. Measurement of similarity in images which differ in terms of brightness and contrast can be solved with the normalized covariance measure (NCM) [1]. The NCM computes covariance between voxel intensities in two image blocks and divides it by a geometric mean of intensity variances in the blocks. The most complex situation is to compare images of differing modalities. This problem is often solved by applying the mutual information (MI) measure [3], which compares information entropies in the compared image blocks. In NCM and MI, the more similar the image blocks, the higher is the value of the measure.

The definitions of the MAD and NCM functions are given in equations (8) and (9). They define relation between the block a of voxels in image I_A and the similar block b with the center at coordinates (x_b, y_b, z_b) in image I_B. The R parameter defines a so-called radius of the block.

$$MAD_a(x_b, y_b, z_b) = \frac{1}{(2R+1)^3} \sum_{i=-R}^{R} \sum_{j=-R}^{R} \sum_{k=-R}^{R} |I_A(x_a+i, y_a+j, z_a+k) - I_B(x_b+i, y_b+j, z_b+k)| \quad (8)$$

$$NCM_a(x_b, y_b, z_b) = \frac{\sum_{i=-R}^{R} \sum_{j=-R}^{R} \sum_{k=-R}^{R} (I_A(x_a+i,y_a+j,z_a+k)-imu_a)(I_B(x_b+i,y_b+j,z_b+k)-imu_b)}{\sqrt{\sum_{i=-R}^{R} \sum_{j=-R}^{R} \sum_{k=-R}^{R} (I_A(x_a+i,y_a+j,z_a+k)-imu_a)^2} \sqrt{\sum_{i=-R}^{R} \sum_{j=-R}^{R} \sum_{k=-R}^{R} (I_B(x_b+i,y_b+j,z_b+k)-imu_b)^2}} \quad (9)$$

In some solutions the image I_A is divided into blocks of the same sizes organized in a regular raster. Then, for every block its counterpart is searched in the image I_B. This approach is often criticized since not all the blocks from I_A can be uniquely matched with specific blocks in I_B. The example of such situation may be a block presenting a part of homogeneous background region, which will match any location of the similar background in the other image. To solve this problem, the concept of feature points was introduced. Only the blocks presenting specific and unique content of image I_A are matched with the image I_B. This means, the blocks presenting homogeneous regions or recurring content should be excluded from this process. Having this in mind, the regularization of three-dimensional image registration should support the concept of feature point matching and also should enable application of alternative similarity measures.

4 Elastic registration

There are two components considered in image registration, the first is derived from the images and involves image similarity measures, the second is the regularization term and it is derived from the transformation of paired points. The way the two components are combined was inspired by a concept of deformable parametric models or meshes [6] (Fig. 2). This concept defines an energy of the model as a sum of the two components. The energy derived from the image content (from block matching) is usually called an external, and the regularization term is referred as an internal component. The model's energy can be defined by (10), where the parameters ρ and ξ control a contribution of both the energy components, and P is the number of all the points (nodes) of the model. The equation contains an error ε defined by (2). However, note that in section 2 the error was adopted as a function of elements of $\mathbf{J}$ and $\mathbf{T}$. Now, the error is used as a function of coordinates of all the nodes, with assumption that the optimal $\mathbf{J}$ and $\mathbf{T}$ were already established. The function M_p is a image block disimilarity function, computed for a block from the image I_A and linked to the point p. The M_p may be represented by the MAD function, or optionally can be equal to inverted or negated NCM or MI functions.

$$E = \sum_{p=1}^{P} (\rho M_p(x, y, z)) + \xi\varepsilon \tag{10}$$

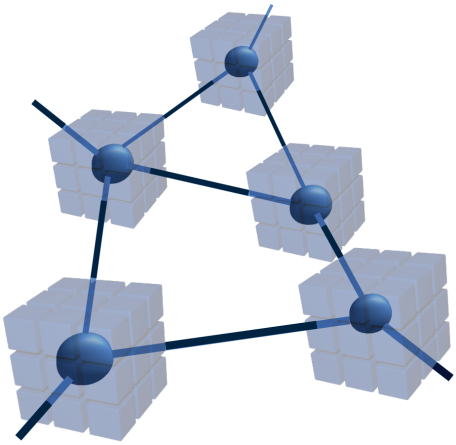

Fig. 2. Conception of deformable mesh. Nodes are linked with each-other to reflect the regularization term and linked with matching blocks to depict the image derived term.

Solving the registration problem requires minimization of the energy E. If we neglect the influence of point p on the $\mathbf{J}$ matrix and on the $\mathbf{T}$ vector, which

is valid if the displacement of the points are small, the problem can be solved independently for individual points. To do this, we apply a gradient-descent optimization (11), where (i) indicates an iteration step and $\mathbf{v}_p = [x_p\ y_p\ z_p]^T$.

$$\mathbf{v}_p^{(i+1)} = \mathbf{v}_p^{(i)} - \rho \nabla M_p(x_p^{(i)}, y_p^{(i)}, z_p^{(i)}) - \xi \nabla \varepsilon_p^{(i)}(x_p^{(i)}, y_p^{(i)}, z_p^{(i)}) \tag{11}$$

$$\nabla \varepsilon_p = \mathbf{v}_p - (\mathbf{J}_p \mathbf{w}_p + \mathbf{T}_p) \tag{12}$$

It must be noted that the coordinates $\mathbf{v}$ change in subsequent iterations. This affects $\mathbf{J}$ and $\mathbf{T}$ elements, which in turn affect a form of the εfunction. Therefore in equation (11) the ε function is indexed with (i) – the iteration number. Moreover, in (12) the ε, $\mathbf{J}$ and $\mathbf{T}$ are indexed with the index p. This means, that the function is computed in different ways for different nodes of the mesh. On the other hand, examination of equation (1) may lead to conclusion that the matrix $\mathbf{J}$ and vector $\mathbf{T}$ are computed the same way for all the nodes in the mesh. This inconsistency requires explanation.

As aforementioned, the affine transformation preserves straight lines, flat planes and keeps their parallelism. Unfortunately this means that the transformation cannot model bending which may appear in some human or animal body structures. Therefore, applying the same matrix $\mathbf{J}$ and vector $\mathbf{T}$ for all the nodes of the mesh would prevent the mesh from bending. To overcome this difficulty the affine transformation is estimated locally within a limited neighborhood of nodes surrounding the selected node p. If the neighborhood is narrow then the bending of the whole mesh is possible. Otherwise, use of wide neighborhoods would counteract bending, and using all the nodes for estimation of the transformation would eventually prevent bending completely.

The procedure for image registration by means of the proposed regularization method is as follows:

1. The feature points are selected in the image to be transformed (I_A – a moving image). The points should be persistently linked with the centers of unique blocks in the image.

2. The mesh is constructed from the points by virtually linking the neighboring points. This step has an impact on ability of the mesh to bend, since narrow neighborhoods enable bending and the wider neighborhoods may restrict bending ability.

3. The ρ and ξ parameters are set to establish contribution of image similarity function term and the regularization term respectively.

4. The transformation matrix and translation vector are computed for every node and its neighborhood. The matrix is decomposed and modified by removal of its selected factors (e.g. removal of the determinant introduces resistance to scaling).

5. For every point, and its persistently linked block, the gradient of image dissimilarity function is computed in the matched (a fixed) image (I_B).

6. New coordinates of every point of the mesh are computed from (11).

7. The steps 4, 5 and 6 are repeated for a given number of iterations. In the final iterations the values of ρ and ξ parameters can be gradually and proportionally reduced.

8. Finally, the moving image shall be transformed from the initial to the final location of the mesh.

5 Results

The registration algorithm was tested on 20 cases of time series showing Dynamic Contrast Enhanced (DCE) MRI images. Each of the time series is made up of 74 three-dimensional images, each of 192×192×30 voxels and with voxel spacing of 2.2×2.2×3 mm. The images present flow of tracer agent into both kidneys, filtration process and outflow of the agent through the renal pelvis. In the subsequent image time frames, brightness of medulla, renal pyramids and renal cortex significantly changes over time. Moreover, the kidneys move in antero-posterior direction due to respiration process, and are slightly squeezed and bent by movement of the diaphragm. Local changes in brrightness and deformations make the registration task nontrivial.

*****TODO: opis wyboru i sposobu łączenia węzłów********** For the 17^{th} frame in every series, the regions of kidneys were manually outlined and the feature points were identified exclusively within the regions. Fig. 3 presents an example mesh of interconnected feature points placed in the three-dimensional region of the kidney. Next, for every series, the registration procedure was performed to adjust the mesh to the remaining frames of the series. Based on the final and the initial form of the mesh, all the frames were transformed to match the image content in the 17^{th} frame. The results were qualitatively assessed by the expert, were recognized as adequate and enabled correct estimation of glomerular filtration rate.

The number of nodes ranged from 550 to 1015 depending on the specific time series. The NCM was used as the image similarity function with a block of 7×7×7 voxels and parameter $\xi = 5.0$. Unimodal transformation was used as a regularization term with $\rho = 0.7$. The registration procedure was split into two stages. At first, 150 iterations with the neighborhood comprising all the nodes of the mesh were executed. In the second stage additional 10 iterations were performed with neighborhoods including nodes located within a range of 12 voxels. The mesh in the first stage behaved semi-rigid, and in the second stage it allowed for bending. This multistage approach was successfully applied in two-dimensional models [6] and proved to be computationally efficient and accurate. The algorithm enabled registration of 4 to 8 images per second on Intel Core i7-4790 360GHz processor.

It must be noted that quantitative assessment of the correspondence in the registered images is difficult since there is no ground truth. Also, an attempt to manually indicate the corresponding points in the images is burdened with a significant error. Therefore, we present the results of registration in a form of

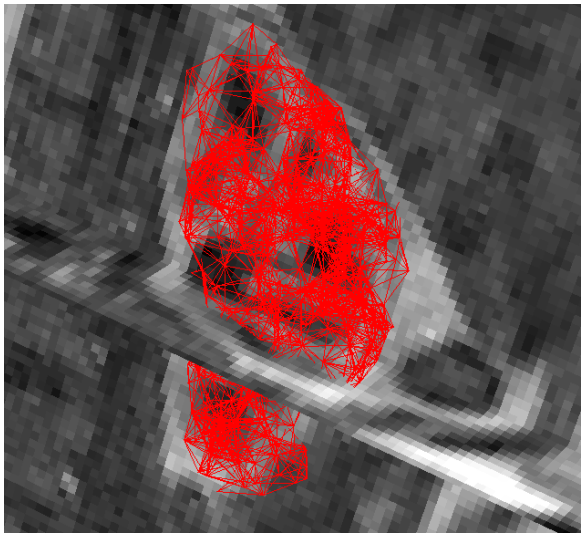

Fig. 3. The example mesh placed in the space of a three-dimensional image.

overlapping images to enable the reader to make his or her own assessment of the method.

Fig. 4 presents selected frames before (LHS) and after (RHS) the registration. The fusion of two co-registered images is presented by means of color components, one of the images is shown in green and the other in magenta. If both the images overlap, the green and magenta contours and structures align to compose a gray-scale pattern. Otherwise the misaligned green or magenta streaks or patches are present. The alignment of the kidney regions after the registration seams accurate. It can be noticed however, that some image fragments outside the kidney volumes, especially the outer contour of the body, are misaligned. This effect was expected and it is acceptable since exclusively the kidney regions were registered, whilst locations of the other image fragments were ignored as irrelevant.

For comparison with a state-of-the-art method, the DCE-MR images were co-registered using b-spline deformable registration algorithm implemented in *Plastimatch* module [4] of *3D Slicer* software. The entire procedure was fully automatic, i.e. no fiducial points were annotated on the registered frames. For each analyzed series, one time-frame was selected as a reference (*fixed*) and all the other frames were registered to it. In *Plastimatch*, image matching can be configured as a multi-stage procedure, where each stage corresponds to a different scale of image content. The first stage performs preliminary alignment based on translation and affine transformation. This step is followed by three b-spline deformable stages varied by the image subsampling rate (in x, y, and z direction) and the grid size. In the experiment we used the following parameter settings: subsampling rate = (4,4,2) grid size = 100 mm (in stage 2), and subsampling

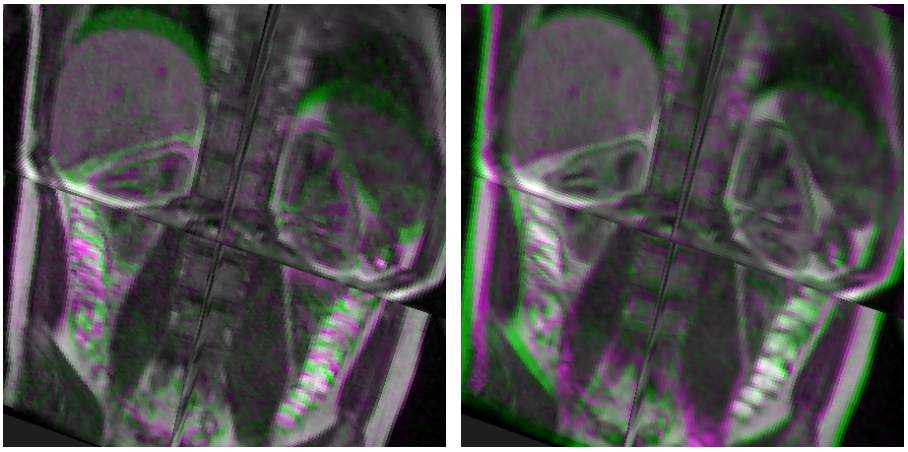

Fig. 4. The images of kidneys before and after the registration

rate = (2,2,1) with grid sizes = 50 and 25 mm (in stages 3 and 4). In each stage, registration was evaluated using the mean squared error criterion. The four registration stages are performed in a sequence and the result of a given step is simultaneously an input to the next one. This procedure provided a registration rate of roughly 1 image per second.

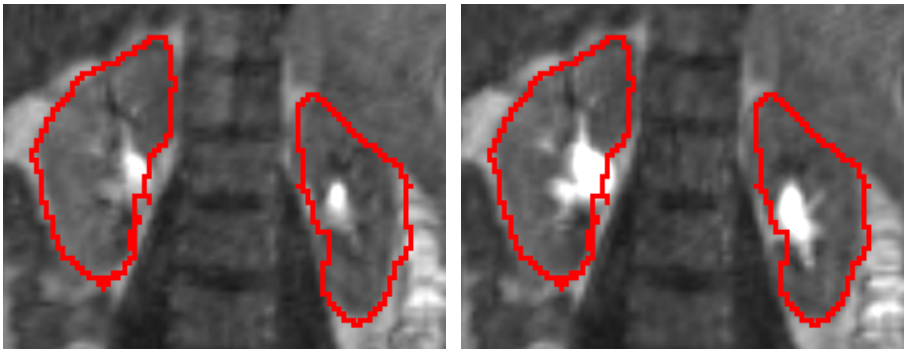

Fig. 5. Comparison of registration results by the reference (LHS) and the proposed (RHS) methods.

Fig. 5 presents frame 63 of the 1st time series. After *Plastimatch* application the shape and size of the white spot of pelvis is much smaller than required. Moreover, the boundary of kidneys are not aligned with the red contour indicating the expected location of the boundary. This means that in the reference algorithm the volume of the whole kidney is inadequate, the boundaries are not properly aligned, and the volume of the pelvis is reduced in an undesirable way.

These unwanted artifacts are not present in the image obtained by the proposed algorithm.

6 Conclusions

The presented algorithm fulfills the needs of medical image registration. It enables to focus the registration on selected regions of interest and on selected feature points. The user has gained control to properly balance the image derived contribution of individual feature points and the contribution of the regularization term, which maintain mutual spatial relations of these points arrangement. The regularization term is based on the locally estimated transformation. The optional choice of affine, Procrustes or unimodal transform make feasible to model various physical properties of tissues, including susceptibility or resistance to stretching, bending or volume changes. The proposed method can be successfully applied to two- or three-dimensional images. Moreover, the algorithm enables arbitrary selection of image similarity function. Therefore, it enables registration of images of the same or varying modalities.

Compared to the reference method, the proposed regularization term and deformable mesh performed more accurate alignment of kidney regions. The position of the kidneys is more stable and the motion effect is completely compensated. The ability to preserve volume has been confirmed in registration of the pelvis region. The proposed solution keeps the shape and the size of the pelvis, whereas the method based on b-splines extremely reduces its volume when the contrast agent is present. The presented results confirmed the ability of the proposed method to correctly register time series of kidney images. The visual assessment of the resulting images confirmed the high accuracy of the registration. The results were found useful for further analysis, specifically for estimation of the glomerular filtration rate.

The further work will focus on quantitative evaluation of the algorithm, its comparison with another state-of-the-art methods, and application to registration of images of various modalities.

Acknowledgement. This work was supported by the Polish National Science Centre Grant 2014/15/B/ST7/05227. The DCE-MR images of kidneys were provided by the Department of Radiology, Haukeland University Hospital, Bergen, Norway. The source codes of the presented program for image registration are available from *https* : *//gitlab.com/piotr.szczypinski/deformowalne*.

References

1. Chen, F., Loizou, P.C.: Analysis of a simplified normalized covariance measure based on binary weighting functions for predicting the intelligibility of noise-suppressed speech. The Journal of the Acoustical Society of America **128**(6), 3715–3723 (2010)
2. Maintz, J.A., Viergever, M.A.: A survey of medical image registration. Medical image analysis **2**(1), 1–36 (1998)

3. Pluim, J.P., Maintz, J.A., Viergever, M.A.: Mutual-information-based registration of medical images: a survey. IEEE transactions on medical imaging **22**(8), 986–1004 (2003)
4. Sharp, G.C., Li, R., Wolfgang, J., Chen, G., Peroni, M., Spadea, M.F., Mori, S., Zhang, J., Shackleford, J., Kandasamy, N.: Plastimatch-an open source software suite for radiotherapy image processing. In: Proceedings ICCR, Amsterdam, Netherlands (2010)
5. Späth, H.: Fitting affine and orthogonal transformations between two sets of points. Mathematical Communications **9**(1), 27–34 (2004)
6. Szczypiński, P.M., Sriram, R.D., Sriram, P.V., Reddy, D.N.: A model of deformable rings for interpretation of wireless capsule endoscopic videos. Medical Image Analysis **13**(2), 312–324 (2009)