# A multi-stage registration method using texture features

Andreja Jarc[1,2,*], Janez Perš[2], Stanislav Kovačič[2]

[1]Sipronika d.o.o., Trzaska 2, SI-1000 Ljubljana, Slovenia
*(corr. author contact: andreja.jarc@sipronika.si)
[2]University of Ljubljana, Faculty of Electrical Engineering, Trzaska 25, SI-1000 Ljubljana, Slovenia

We present a novel, multi-stage registration method based on Laws' texture features. In general, a large number of texture features may be extracted from the original intensity images. For each of the texture features, a criterion function that measures the similarity between the images may be derived. The proposed registration method consists of two major steps. In the first step, a data set of images with the corresponding gold standard is used. In this step, the selection and ranking of the texture features for registration is made. The selection and ranking of the features is based on their robustness, accuracy and capture range. The selected features are then entered in the second step, where the actual registration is performed using a sequence of registration stages. Our method is based on the selection of the most robust feature for the first registration stage and the selection of accurate feature(s) for the subsequent stages. The texture features are daisy-chained so that the accuracy of the previous feature is sufficient for the capture range of the next feature. We tested our method on 11 2D image pairs containing Digital Reconstructed Radiographs (DRR) and Electron Portal Imaging (EPI) modalities, which were difficult to register using intensity features alone. With our method we have successfully registered 75% of the initial displacements, ranging from 5 to 7.5 mm, with the target-registration error below 3 mm, whereas the traditional intensity-based approach delivered only 15% successfully registered cases.

## 1. Introduction

Clinical diagnosis, as well as therapy planning and evaluation, rely increasingly on information integration, using multiple images of different modalities. For example, in radiation-therapy planning a CT (computer tomography) scan is needed for the dose-distribution calculations, while the contours of the target lesion are often best outlined on an MRI (magnetic resonance image) [1]. One of the fundamental tools of information integration is image registration. The detailed classifications of registration techniques applied to medical images have been reviewed in a number of surveys [2,3,4,5,6,7].

In general, image registration is implemented as an optimization process for finding the transformation parameters that maximize or minimize a criterion function (CF) that measures the similarity between the images. There are several different methods for computing a CF. The methods can be classified into two categories according to the features used: geometrical-feature-based and intensity-based methods [8].

Numerous examples in the literature show promising results for registration based on geometrical features [1,9,10,11]. However, geometrical features are obtained through a segmentation step, before the actual registration is performed. Segmentation is, by itself, already a complex and tedious task, and therefore the success of geometrical-feature-based registration is mainly dependent on the success of the segmentation step. Moreover, the

precise segmentation of anatomical geometrical features is generally difficult to automate.

Intensity features [12], in contrast to geometrical features, do not require segmentation, and therefore intensity-based registration can be fully automated. However, intensity features do not explicitly show structural information, which may be crucial for the registration of some image modalities. In those cases, intensity features alone may not be sufficient for the registration. This may be due to image artifacts, low image quality, a 2D projection of 3D data or a low image overlap [8,13,14,15,16].

An alternative approach to registration is texture features. Texture features are frequently used in computer vision for image analysis, synthesis and recognition. Texture-feature-based image registration is rarely used, but, according to the literature, some experiments show that it delivers promising results in the case of poor-quality ultrasound images [13,14]. A special class of texture-based approaches described in the literature [16,17,18,19,20,21] aims to combine the image gradient and edge information with the intensity.

The appeal of texture features is that they explicitly show the local structures contained in an image, in a similar way to geometrical features. On the other hand, they can easily be automated, in a similar way to intensity features. Nevertheless, the selection of the appropriate texture features is crucial for a successful registration. Different texture features have significantly different properties, and therefore the criterion functions based on them may have very different properties. In particular, when used for registration, they may differ in terms of robustness, accuracy and capture range, and therefore a tool for the selection of the features is needed.

In this paper we propose a method consisting of an algorithm for the selection of the texture features and an accompanying algorithm for their use. Our method is based on an analysis of the features on a data set of images with the corresponding gold standards (GSs). We expect that such a procedure would have to be done for each class of registration tasks, as the feature selection is valid only for a certain combination of modalities and, probably, only for certain anatomical structures. On the other hand, texture features enable us to do the registration in a similar way as with intensity features; however, they have some potential with image modalities, where the intensity features fail.

To select the appropriate texture features for registration we propose an extended protocol for the evaluation of the similarity measures for rigid registration, originally devised by Škerl et al. [22]. The evaluation protocol assesses the quality of the similarity measure used in a specific registration problem prior to the registration. This is done by evaluating the behavior of the similarity measure for a set of simulated transformations. The evaluation of a similarity measure includes the three main parameters: accuracy, robustness and capture range.

This evaluation gives us the basic information needed for the selection of the texture features, which is done once on a representative data set of image pairs, with the corresponding gold standard, prior to the registration. Evaluation also provides feedback on the appropriateness of the selected features, and, indirectly, provides an estimate of how well such registration will perform, when done on the images with similar characteristics.

Following the feature selection, the actual registration is performed using a sequence of stages, where different pre-selected texture features are used. Our method is based on the use of the most robust feature in the first stage and the use of more accurate feature(s) in the subsequent stages. In short, the texture features are daisy-chained, so that the accuracy of the previous feature is sufficient for the capture range of the next feature.

The experiments were performed on 11 pairs of DRR (Digital Reconstructed Radiograph) and EPI (Electron Portal Imaging) images of the pelvis. DRR images are usually obtained in advance of radiotherapy treatment; they are used for therapy planning and are

essentially 2D projections of 3D CT volumes. Therefore, their basic content is the same as the content of CT images. On the other hand, the EPI images are obtained during the treatment. The same radiation source is used for image acquisition as for the therapy itself. First, a low dose is used to acquire an image, which is then checked against the DRR image for proper patient alignment. If the patient is positioned within the therapy-specific tolerances, the radiotherapy is performed, this time using a more powerful beam. Otherwise, the position of the patient is corrected. Portal imaging is an important method for measuring and documenting the extent of the geometric treatment's accuracy [23].

The main aim of the radiation therapy is to deliver a prescribed radiation dose as accurately as possible to a tumor region while minimizing the dose's effect on the neighboring normal tissues. Therefore, proper patient positioning is extremely important and DRR/EPI registration could be of significant help in this task. The imaging setup for the acquisition of EPI images is specialized for the therapy, not for the image-based diagnosis. Due to the specifics of the EPI devices (mega-volt beam) the contrast of the EPI images is low, as the different tissues exhibit only small differences in their absorption coefficients. Therefore, anatomical structures are not clearly visible, and some are not visible at all, which makes some of the frequently used registration methods (e.g., intensity-based registration) inefficient. Texture features, on the other hand, extract local structural information and may provide better results on such images.

The remainder of the paper is organized as follows. First, the evaluation protocol for the similarity measures is described, along with our modifications and extensions. Then, the texture features used in our tests are introduced, followed by an explanation of the multi-stage registration algorithm. Next, the data set and the experimental design are presented, and the results of the tests are described and discussed. Finally, we present some conclusions regarding the proposed method.

# 2. Materials and methods

During the selection step we need to obtain information about the appropriateness of each of the texture features for a given registration problem. For this task we extended the protocol for the evaluation of the similarity measures for rigid registration [22]. The evaluation protocol is applied to a data set of image pairs for which we have a gold standard. It has been tested for various multi-modal rigid-registration tasks [22,24,25]. The output of the evaluation protocol is the parameters that measure the appropriateness of the criterion functions for a given registration task. In the case of texture-based registration, the texture feature itself forms a part of the criterion function. Therefore, the evaluation protocol is also used to evaluate the texture features.

The protocol is briefly explained in the following section however, for details the reader is referred to [22].

## *2.1. Evaluation protocol*

The evaluation procedure is as follows. The continuous $K$-dimensional space of the transformation parameters is first normalized, so that equal changes to each of the parameters in the normalized parametrical space produce the same mean voxel shift. $K$ depends on the transformation model used, for example, $K=6$ for the 3D, and $K=3$ for the 2D rigid transformation models, respectively. Next, the normalized $K$-dimensional space is "probed" by $N$ randomly selected sampling lines. The lines are positioned in such a way that their intersection points with a hyper-sphere with radius $R$ are uniformly distributed on the surface of the hyper-sphere. All the sampling lines intersect in the gold-standard (GS) transformation, which corresponds to the aligned position of two images. Each sampling line is subsequently

sampled by $M$ equidistant points and the step size between the points is defined as $2R/M$. Let us denote $X_0$ as the origin or GS transformation and $X_{n,m}$ as one of the sampled points. Then each of the $X_{n,m}$ represents a $K$-dimensional vector of the transformation parameters (see Fig.1(The figure is reproduced from [22])).

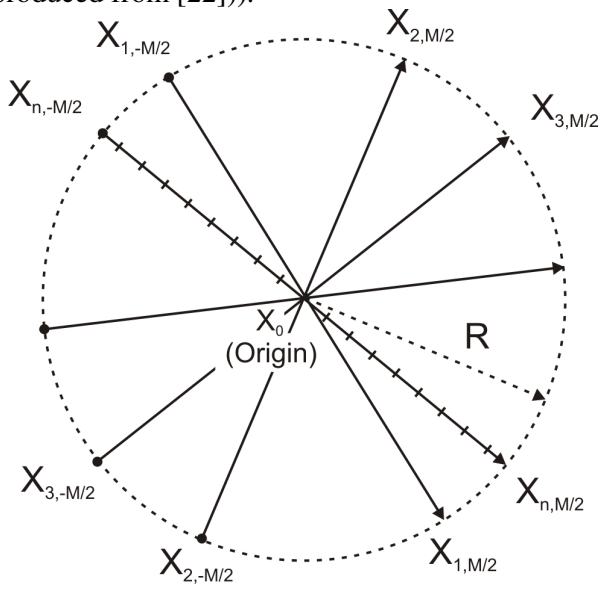





**Fig. 1.** 2D parametrical space, sampled by $N$ lines and $M$ points per line. The maximum displacement from the GS is denoted by $R$, which is the radius of the $K$-dimensional hyper-sphere. $M$ and $R$ define the step size between sampling points: $2R/M$.

The normalization parameters used to generate the sampling lines for the image pairs used in our experiments are listed in Table 1. For a detailed explanation of the parameters, see [22].

**Table 1.** Image sizes, pixel sizes, translation and rotation units of normalized parametrical space, radius $R$, number of sampling lines $N$, number of points along a line $M$ and step size between two sampling points $\delta$.

| Image pair | Image size (mm) | | Pixel size (mm) | | Unit (mm) | Unit (rad) | R(mm) | N | M | δ(mm) |
|---|---|---|---|---|---|---|---|---|---|---|
| | X | Y | X | Y | | | | | | |
| 01 | 203 | 170 | 0.52 | 0.52 | 17.0 | 0.13 | 51.0 | 10 | 400 | 0.26 |
| 02 | 205 | 179 | 0.52 | 0.52 | 17.9 | 0.13 | 53.7 | 10 | 400 | 0.27 |
| 03 | 258 | 190 | 0.52 | 0.52 | 19.0 | 0.12 | 57.0 | 10 | 400 | 0.29 |
| 04 | 203 | 151 | 0.52 | 0.52 | 15.1 | 0.12 | 45.3 | 10 | 400 | 0.23 |
| 05 | 246 | 140 | 0.52 | 0.52 | 14.0 | 0.10 | 42.0 | 10 | 400 | 0.21 |
| 06 | 194 | 165 | 0.52 | 0.52 | 16.5 | 0.13 | 49.5 | 10 | 400 | 0.25 |
| 07 | 254 | 173 | 0.52 | 0.52 | 17.3 | 0.11 | 51.9 | 10 | 400 | 0.26 |
| 08 | 201 | 162 | 0.52 | 0.52 | 16.2 | 0.13 | 48.6 | 10 | 400 | 0.24 |
| 09 | 206 | 123 | 0.52 | 0.52 | 12.3 | 0.10 | 36.9 | 10 | 400 | 0.18 |
| 10 | 248 | 188 | 0.52 | 0.52 | 18.8 | 0.12 | 56.4 | 10 | 400 | 0.28 |
| 11 | 195 | 107 | 0.52 | 0.52 | 10.7 | 0.10 | 32.1 | 10 | 400 | 0.16 |

The original evaluation protocol was modified for a more efficient evaluation of a large number of criterion functions, which is the case when texture features are evaluated. The modifications are detailed in [26]; here we provide just a brief summary. The modification was implemented using Halton quasi-random sampling instead of pseudo-random sampling. The results obtained using quasi-random sampling were compared to the results of pseudo-random sampling, which is used in the original version of the protocol. We found that with the use of quasi-random sampling the number of sampling lines for the 2D registration task can be significantly lower to obtain results that are comparable to the results obtained using the original protocol.

In the evaluation protocol each sampling line provides a transformation profile for a similarity measure, which in our case includes a chosen texture feature. To observe the behavior of such a similarity measure, three parameters are used: the accuracy (*ACC*), the risk of non-convergence (*RON*) and the capture range (*CR*). The accuracy, as defined in [22], is the root-mean-square distance between the maximum value $X_{n,max}$ of the criterion function and the origin $X_0$ on each of the $n$ sampling lines: $n = 1,2,\ldots N$. For the image pair $i$, the $ACC^i$ is calculated as:

$$ACC^i = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left\|X_{n,\max} - X_0\right\|^2}. \tag{1}$$

The risk of non-convergence *RON(r)* is defined as the average of the positive gradients $d_{n,m}$ within a distance $r$ from each of the $N$ global maxima. For the image pair $i$, the $RON^i$ is calculated as:

$$RON(r)^i = \frac{1}{2rN}\sum_{n=1}^{N}\sum_{m=\max-k}^{\max+k} d_{n,m}. \tag{2}$$

In our case, $r=R$, as specified in Table 1, and therefore we use *RON=RON(R)* throughout the paper.

In addition to the values of $RON^i$ and $ACC^i$, we computed the standard deviations $\sigma_{RON^i}$ and $\sigma_{ACC^i}$ among the sampling lines for each image pair.

The capture range (*CR*) is defined as the smallest of the $N$ distances between the positions of global maxima and the closest local maximum $X_{n,loc}$ along each line. For the image pair $i$, the $CR^i$ is calculated as follows:

$$CR^i = \min_n\left(\left\|X_{n,\max} - X_{n,loc}\right\|\right) \tag{3}$$

To assess the appropriateness of a texture feature for the registration of a given image pair we augmented the original evaluation protocol with a definition of the *quality* (*Q*) of a feature. There are two basic quality indicators for each feature. First, it can have a low risk of non-convergence. This risk is manifested through a low *RON* value and a low deviation of the *RON* values among the sampling lines. Therefore, we define the $Q^i_{RON}$ for the image pair $i$ as:

$$Q^i_{RON} = \left(\alpha_1\overline{RON^i} + \alpha_2\frac{\sigma_{RON^i}}{\overline{RON^i}}\right)^{-1}, \tag{4}$$

where $\overline{RON^i}$ is the mean value of $RON^i$, calculated across all the sampling lines for the image pair $i$, $\sigma_{RON^i}$ is the standard deviation of $RON^i$, and $\alpha_1$ and $\alpha_2$ are the weights, which were set to 1 in our experiments. Essentially, $Q_{RON}$ contains both the actual risk of non-

convergence and its consistency among the sampling lines. By adjusting $\alpha_1$ and $\alpha_2$ we were able to place greater emphasis on either of those two components.

The second quality of a texture feature is related to its ability to register the images accurately. This ability is manifested through low $ACC$ values and a low deviation among the sampling lines. Therefore, we define the $Q_{ACC}$ for the image pair $i$ as:

$$Q_{ACC}^{i} = (\beta_1 \overline{ACC^i} + \beta_2 \frac{\sigma_{ACC^i}}{\overline{ACC^i}})^{-1}, \tag{5}$$

where $\overline{ACC^i}$ is the mean value of $ACC^i$, calculated across all the sampling lines for the image pair $i$, $\sigma_{ACC^i}$ is the standard deviation of $ACC^i$, and $\beta_1$ and $\beta_2$ are the weights, which were set to 1 in our experiments. $\beta_1$ and $\beta_2$ have the same effect on $Q_{ACC}$ as $\alpha_1$ and $\alpha_2$ have on $Q_{RON}$.

In the rest of the paper we use the following terminology: the feature with a small value of *RON* is said to be *robust*; the feature with a small value of *ACC* is said to be *accurate*.

## *2.2. Texture features used for registration*

For texture-feature-based registration we first extract the texture features from the images. In our work we use multi-scale Laws' texture coefficients, yielding in total 48 different texture-feature images. Each texture-feature image can be registered in the same way as intensity images and the intensity image can be treated simply as one additional feature image.

The same texture features were extracted from both modalities of each image pair. In the registration step we considered only the registration of the same texture features.

The texture features used in our work are all rotationally invariant. We would expect that texture features that depend on the angle would perform poorly, since we aim to register both the spatial and the rotational misalignments. This was also confirmed in our preliminary tests, leading us to the choice of rotationally invariant Laws' texture features.

### 2.2.1. Laws' texture features

The 2D Laws' filter masks [27] are derived from a set of 1D filters, where each of the 1D filters extracts a certain texture feature from an image. These features are: level (L), edge (E), spot (S), wave (W) and ripple (R). For example, the 1D 5-sample filter masks are:

L5=[1 4 6 4 1]

E5=[-1 -2 0 2 1]

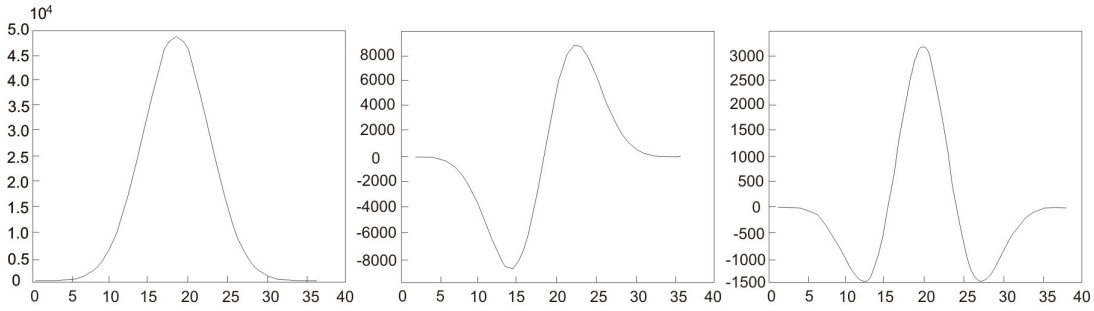
S5=[-1 0 2 0 -1]

W5=[-1 2 0 -2 1]

R5=[1 -4 6 -4 1].

According to Laws, the most useful texture features are obtained from a combination of E-edge, S-spot and L-level filter masks [28]. On the basis of this we limited ourselves to those texture features. By convolving these filter masks with each other, a set of symmetric and anti-symmetric filters are obtained. Each of the E_L, S_E and S_L are added to their

transposed pairs L_E, E_S and L_S, respectively, thus generating rotationally invariant texture features. To deal with structures that have different anatomical sizes, we systematically generate texture features for the sizes $n$=1,5,10 and 20 mm by re-sampling the original 5-sample filter masks *E5*, *L5* and *S5* to a window size $n$. In this way 12 different texture features are obtained. An example of the 1D filter masks 'L ', 'E' and 'S' of size 20 mm is given in (Fig. 2).




a) 'L' filter mask b) 'E' filter mask c) 'S' filter mask

**Fig. 2.** a) Filter mask 'L', the pixel size in the $x$ direction was 0.56 mm, therefore 20mm/0.56 mm = 36 units (pixels). The mask was obtained with the interpolation of the original 'L' mask of size 5 units. b) Filter mask 'E' of size 36 units. The mask was obtained with the interpolation of the original 'E' mask of size 5 units. c) Filter mask 'S' of size 36 units. The mask was obtained with interpolation of the original 'S' mask of size 5 units.

Additionally, we apply a different amount of local Gaussian smoothing to each of the texture features. We use the Gaussian kernel of size 1, 5, 10 and 20 mm, where the width of the filters was $w=6\sigma$. A combination of 12 texture features and 4 different smoothing kernels yields a final number of 48 texture-energy features. The combinations of features are shown in Table 2. Finally, the extreme 2% of the values of each texture-energy image are saturated, to account for the potential intensity artifacts, and the rest are scaled and quantized to 8 bits.

**Table 2.** Filter combinations and their sizes. There were 48 combinations, yielding 48 Laws' texture features.

| Combinations of Laws filter masks | The size $n$ of Laws filter masks [mm] | The size $w$ of Gaussian kernel [mm] |
|---|---|---|
| $E_nL_n$_$L_nE_n$ | 1,5,10 and 20 | 1,5,10 and 20 |
| $S_nL_n$_$L_nS_n$ | 1,5,10 and 20 | 1,5,10 and 20 |
| $E_nS_n$_$S_nE_n$ | 1,5,10 and 20 | 1,5,10 and 20 |

## *2.3. Registration*

The core of our registration procedure, for both the texture-based and intensity-based registrations, is based on Powell's [29] optimization method, which aims to maximize the CF, which in turn measures the similarity between a pair of images for a particular set of transformation parameters. We use a rigid-transformation model with three degrees of freedom (translations in the $x$ and $y$ directions and rotation). The optimization method adjusts

the transformation parameters until the value of a criterion function does not increase more than a pre-defined tolerance value.

We chose mutual information (MI) [12] as a metric to measure the similarity between the images:

$$MI(A,B) = H(A) + H(B) - H(A,B) \tag{6}$$

where $H(A)$ and $H(B)$ are the Shannon marginal entropies of the image features for both of the images and $H(A,B)$ is their joint entropy.

MI is computed based on histograms of the images involved and 2D joint histograms obtained using a partial volume interpolation.

Due to the relatively small dimensions of the input images (see Table 1) the 2D joint histogram is relatively sparse. Therefore, the probability distributions were estimated using the Parzen kernel estimator. The Parzen kernel was applied to the 2D joint histograms, quantized into 256x256 bins. Each joint histogram was created by adding an $(i, j)$ point for every pair of corresponding pixels in the overlap region, where $i$ was the pixel value in the reference image, and $j$ was the interpolated pixel value of the floating image.

## *2.4. Multi-stage registration method*

The proposed multi-stage registration algorithm is made up of two major steps. The first step is the selection of texture features on a training data set of image pairs with the gold standard. The second step is the actual registration, which is done in multiple stages. The choice of features for each of the stages depends on the results of the selection step.

### 2.4.1. Selection step

In the selection step the texture features are extracted from all the images from the training data set. Next, the extended evaluation protocol (Section 2.1) is used to check which features are the most appropriate for a given registration task. It is assumed that the images from the training data set are representative for the given registration problem. The output of the selection step is a list of features sorted by their qualities $Q_{RON}$ and $Q_{ACC}$. The selection step is summarized in Algorithm 1.

---

Algorithm 1: Selection step

---

**Input:** Training data set with provided *GS*
**Output:** Texture feature 0, $ACC_0$, $CR_0$, texture feature 1, $ACC_1$,... texture feature $n$, $ACC_n$
1: Define a bank of texture features that are to be considered for texture-based registration.
2: **for** Each image pair $i$ from the data set **do**
3: Extract all texture features.
4: Use modified evaluation protocol for similarity measures to obtain $ACC^i$ Eq.(1), $RON^i$ Eq.(2) and $CR^i$ Eq.(3) values and their standard deviations.
5: Evaluate the quality of each texture feature, based on $RON^i$, $ACC^i$ and their standard deviations, using Eq.(4) and Eq.(5).
6: Sort the quality values for $Q^i_{RON}$ and $Q^i_{ACC}$ in descending order.
7: For both $Q^i_{RON}$ and $Q^i_{ACC}$ establish four quality classes, corresponding to four quartiles.
8: **end for**

9: **for all** Texture features **do**
10: Count the occurrences of a texture feature in each of the quality quartiles across all image pairs, separately for $Q^i_{RON}$ and $Q^i_{ACC}$.
11: Calculate *ACC*, *RON* and *CR* for the texture feature as mean values of $ACC^i$, $RON^i$ and $CR^i$ across image pairs, respectively.
12: **end for**
13: Choose the most robust texture feature: the one which most frequently appears in the highest-quality quartile, according to $Q^i_{RON}$. The estimation of the largest displacement that the registration will be able to handle is given by *CR* of the chosen feature.
14: **if** *ACC* of selected feature is satisfactory **then**
15: One-stage registration using only the most robust feature will be sufficient.
16: Exit the selection algorithm.
17: **end if**
18: Let $ACC_0$ be the value of *ACC* and $CR_0$ the value of *CR* for the chosen feature.
19: Let $n=1$.
20: **repeat**
21: Find the most accurate feature: the one which most frequently appears in the highest-quality quartile according to $Q^i_{ACC}$.
22: **if** $ACC_n < ACC_{n-1}$ and $CR_n > ACC_{n-1}+ACC_n$ **then**
23: The found feature is the feature of the $n$-th registration stage.
24: Let $ACC_n$ be the *ACC* of the chosen feature.
25: Let $n=n+1$.
26: **else**
27: Mark this feature as a-priori inappropriate for registration at the $n$-th stage.
28: **end if**
29: **until** The texture bank is exhausted or the $ACC_n$ is satisfactory.

---

The algorithm, as formulated, allows either single-stage or multi-stage registration. If the accuracy of the most robust texture feature is adequate, the algorithm will recommend single-stage registration. If the accuracy after the first stage is inadequate, the algorithm will try to find the most suitable texture feature for the next stage. The first condition in line 22 ensures that the added stage will register images with a higher expected accuracy, while the second condition in line 22 ensures that the stages will be added in a daisy-chain manner. Daisy-chaining requires that the feature at the stage $n$ has a capture range of at least $ACC_{n-1} + ACC_n$, as shown in Fig. 3. In this way we can expect that the registration will not break during the transition between stages.

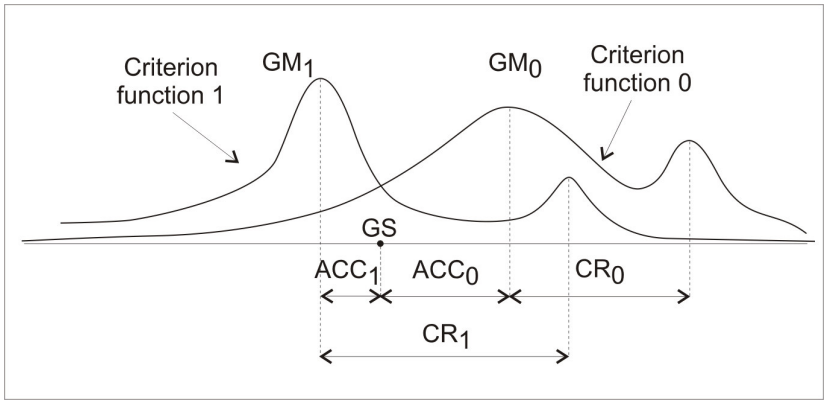



**Fig. 3.** Schematic representation of Algorithm 1 for a two-stage case. Two criterion functions, corresponding to the robust and accurate texture features, respectively, are shown. GS denotes the gold-standard position. $CR_0$ and $CR_1$ denote the capture ranges of the robust and accurate features, respectively. $GM_0$ and $GM_1$ denote the global maxima of the robust and accurate features, respectively.

Note that Algorithm 1 is used only once when a new registration task is encountered. For example, a new clinical problem, new image modalities, etc. Algorithm 1 assumes that we have obtained values of $Q^i_{RON}$ and $Q^i_{ACC}$ for each image pair, which in turn requires that the full evaluation protocol, as defined in Section 2.1, is performed on a supplied image data set with the gold standard available. It is obvious that this is not a task that is performed frequently.

Additionally, as a side effect of calculating $Q_{RON}$ and $Q_{ACC}$ for every texture feature and every image pair, we can estimate how successful the final registration is expected to be. If we examine $Q_{RON}$ and $Q_{ACC}$ for the selected features, we can predict on which images from the data set the registration will fail. This can also serve as a guideline as to how representative is the available data set for the given class of registration problems. If there is a large number of potential failures, then the data set is either inconsistent and therefore poorly represents the registration task, or the number of texture features is too small and does not include features that would be appropriate for the given class of registration problems.

2.4.2. Registration step

The registration step is performed whenever a new pair of images for registration is obtained. The guidelines for the registration are set in the selection step by Algorithm 1 and the registration algorithm follows those steps to obtain the final registration results. The registration step is presented in Algorithm 2.

Algorithm 2: Registration step

**Input:** Texture feature 0, $ACC_0$, $CR_0$, texture feature 1, $ACC_1$,... texture feature $n$, $ACC_n$
1: Perform the registration with the texture feature 0 and constrain the optimization method with $CR_0$.
2: **for** $n$=1 to the number of available features **do**
3: Set the initial transformation of the registration to the resulting transformation of the stage $n$-1.

4: Perform the registration with the texture feature $n$ and constrain the optimization method with $ACC_{n-1}$+$ACC_n$.

5: **end for**

The registration is schematically shown in Fig. 4.

The registration algorithm assumes that the features chosen by the selection step have been extracted from the image pair to be registered. It also assumes that in a selection step the estimates of $ACC$ and $CR$ for those features have been obtained and stored. Those estimates are used for constraining the optimization at each of the registration stages. Each registration stage starts with the transformation from the preceding stage, and the optimization is constrained to minimize the risk of non-convergence, especially when less-robust features on the higher stages are used.

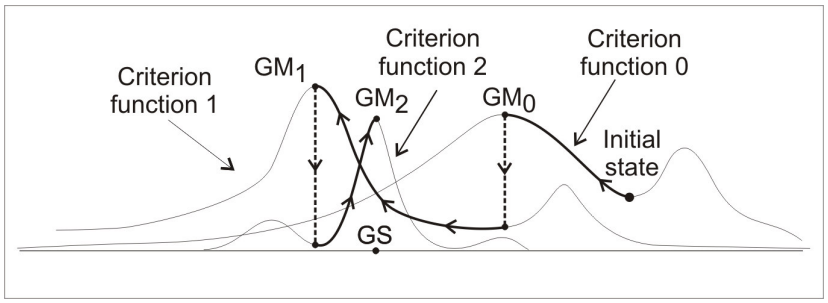
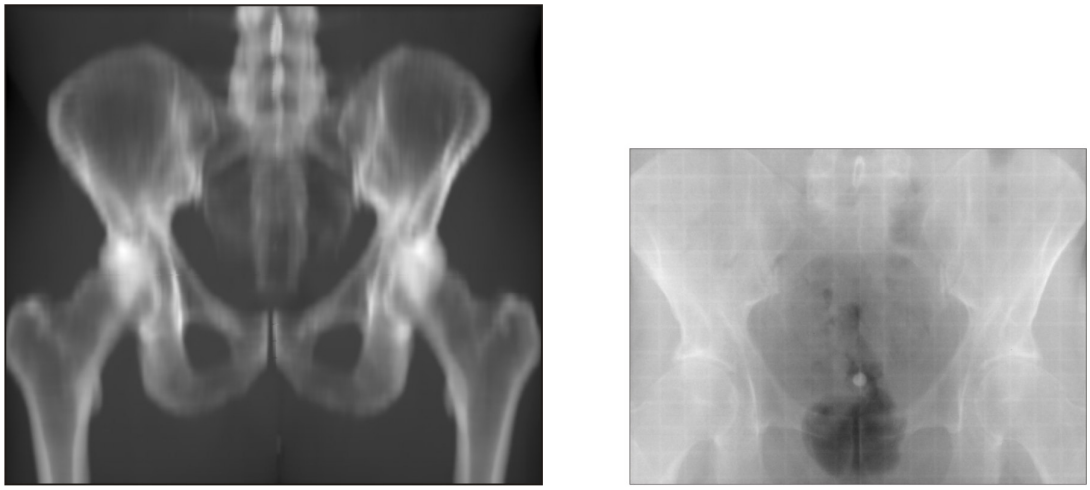



**Fig. 4.** Schematic representation of Algorithm 2 for a three-stage 1D registration case. In the first stage the robust feature (criterion function 0) is used. After reaching the global maximum of the robust feature (GM0) registration proceeds toward the gold standard (GS) using the second-stage feature (criterion function 1), which is more accurate, but with a shorter capture range than the robust feature. Finally, the registration follows the second accurate feature (criterion function 2) to travel from the global maximum of the second feature (GM1) to the global maximum of the third feature (GM2), which is sufficiently close to our gold-standard (GS).

## *2.5. Experimental data set*

The experimental data set consisted of 11 pairs of DRR and EPI images of the pelvis. Examples of both modalities can be seen in Fig. 5.



a) DRR image of the pelvis b) EPI image of the pelvis

**Fig. 5.** An example of one of the intensity image pairs. a) The reference image of resolution 582 x 517 pixels of size 0.56 x 0.56 mm. b) The floating image of resolution 495 x 364 pixels of size 0.52 x 0.52 mm.

Several attempts to register EPI and DRR images using just intensity information have been published [30,31,32,33]. The authors report a successful registration by applying intensity-based methods. However, in our studies we were unable to obtain satisfactory results from our data set using just intensity information. A visual inspection of our EPI images revealed that our images, obtained by radiologists during actual therapy, had markedly lower contrast than in some of the related studies [32]. While we are unsure about the reasons for the lower contrast of our images in comparison to some other studies, it may be because our images originate from clinical practice, where image quality may be sacrificed to satisfy other criteria. However, since both DRR and EPI images depict the same anatomy, it was reasonable to expect [8,15] that the texture-based registration would perform better than the intensity-only registration.

The images used for the tests were initially roughly aligned. The initial image alignment was achieved using three lasers (sagittal, coronal and axial) to mark the patient's reference coordinates [34]. To obtain the gold-standard (GS) alignment needed to evaluate the proposed registration method, the following procedure was used.

Five radiotherapy experts were asked to identify at least five corresponding points for each of the 11 image pairs. These correspondences were used to estimate the transformation parameters between both the images of an image pair. If we found that, after the transformation was estimated, the target registration error (TRE) Eq.(7) of the manually placed points was larger than 3 mm (clinically motivated bound), the expert was asked to repeat the manual process. $T_{GS}$ in Eq.( 7) was, for this case, an identity matrix. This procedure was repeated until the TRE for each image pair and each expert was below 3 mm. The final mean values of the TRE across all the experts for each image pair are shown in Table 3. The TRE is computed as follows:

$$TRE = \left\| T_{GS} \cdot p_{DRR} - T_{estimated} \cdot p_{EPI} \right\| \tag{7}$$

where $p_{DRR}$ denotes points identified on a reference DRR image and $p_{EPI}$ denotes points identified on a floating EPI image. $T_{GS}$ denotes the gold-standard transformation and $T_{estimated}$ denotes the estimated transformation.

| Image pair | TRE [mm] |
|---|---|
| 01 | 1.7 ± 0.7 |
| 02 | 3.0 ± 1.7 |
| 03 | 1.6 ± 0.7 |
| 04 | 1.7 ± 0.8 |
| 05 | 1.8 ± 1.1 |
| 06 | 2.9 ± 1.4 |
| 07 | 1.6 ± 1.0 |
| 08 | 1.6 ± 0.9 |
| 09 | 1.6 ± 1.1 |
| 10 | 2.1 ± 1.0 |
| 11 | 1.9 ± 0.9 |

**Table 3.** Final TRE across the experts for each image pair. The results are given as average ± standard deviation.

All the correspondences (from all the experts) for each image pair were used to obtain the *master transformation* – a gold-standard transformation ($T_{GS}$) for that particular image pair. Using the gold-standard transformation the images were transformed into an aligned position that was the starting point for all our experimental work. The final tolerance of our GS registration, (2.0 ± 1.0) mm, was obtained by observing the TRE across all the points, images, and experts.

## *2.6. Experimental design*

The registration evaluation experiment was designed as follows. Since our data set contained only 11 image pairs with the corresponding gold standard, we used leave-one-out cross validation. In each of the 11 test runs, one image pair was withheld from the set, leaving 10 "training" image pairs. This reduced data set was then used to run Algorithm 1 to select the texture features for the actual registration. Then, the image that was withheld from the training set was registered using those features. In all cases, Algorithm 1 decided that a two-stage registration was sufficient.

Examples of some of the texture features can be seen in Fig. 6. The texture features shown correspond to the image pair shown in Fig. 5, and were selected when this image pair was withheld from the training data set.

The parameters $ACC^i$, $RON^i$ and $CR^i$ were obtained using the evaluation protocol, as defined in Section 2.1. The values of $Q^i_{RON}$ and $Q^i_{ACC}$ were derived from the obtained parameters, using Eq.( 4) and Eq.( 5), respectively.

After each of the image pairs was registered we measured the success rate of the registration (SR%). This success rate was defined as the percentage of registration transformations that, when applied to the gold-standard points, resulted in a target registration error (TRE) Eq.( 7) lower than the clinically motivated bound, determined by several thresholds.

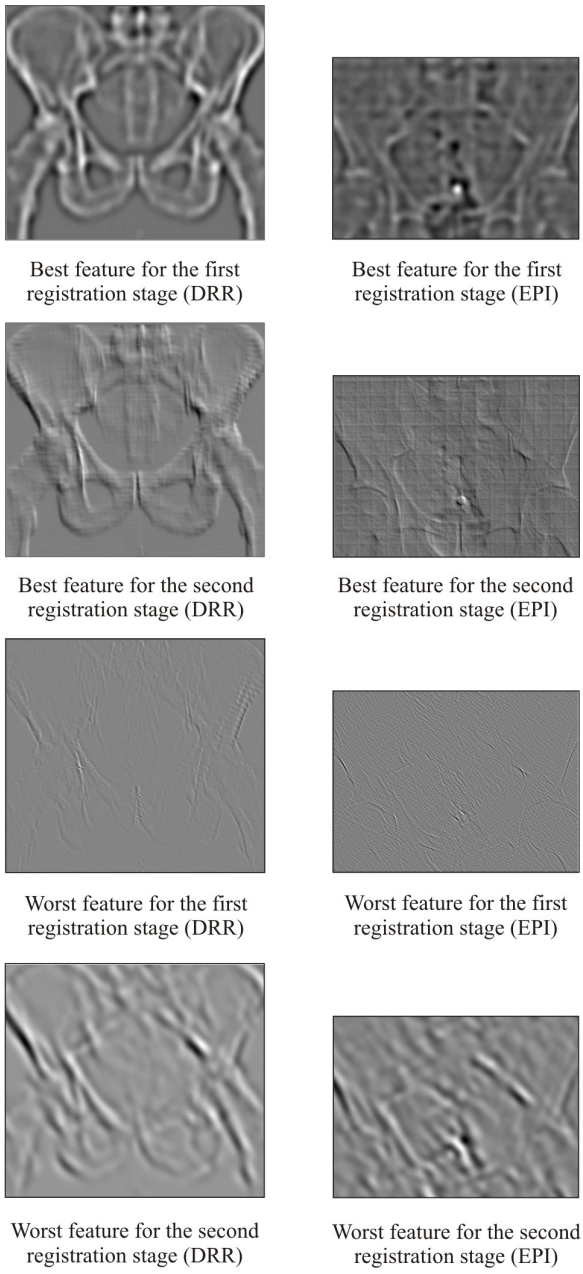



**Fig. 6.** Some examples of texture-feature images. The notions *best* and *worst* are in the context of the test run, when this particular image was absent from the training data set. Images in the left column depict Laws' texture-feature images extracted from a DRR intensity image. The texture-image resolution is 469 x 425 pixels of size 0.56 x 0.56 mm. The dimensions of the texture images are smaller, since we had to crop the images to get rid of the filtering artifacts at the image borders. The images in the right-hand column depict Laws' texture-features extracted from an EPI intensity image. The texture-image resolution is 436 x 309 pixels of size 0.52 x 0.52 mm.

In the first group of experiments, we tested the intensity-based registration. In our framework this was a single-stage registration with the image intensity values in place of the texture features. Based on our initial tests we expected that such a registration would, in

general, not be sufficient.
In the second group of experiments, we tested the proposed method for texture-based registration, as defined by Algorithms 1 and 2. Following the instructions from Algorithm 1, a two-stage approach was used. The chosen features for the first and second stages of the registration are shown in the first and second rows in Fig. 6, respectively. The registration was performed, as specified by Algorithm 2.

### *2.7. Experimental evaluation*

In each experimental group, we performed registrations on a randomly selected set of initial displacements. The initial displacements were obtained as follows. The components of a 3D vector $[x, y, \phi]^T$ were generated randomly, using a pseudo-random generator. The vector, representing a particular displacement, was then classified into one of the five ranges, based on its absolute value. The procedure was repeated until we obtained 20 displacements in each range. The range boundaries were defined as shown in Table 4.

**Table 4.** 100 initial displacements from the GS in the range from 0 to 20 mm.

| Nr. of displacements | Range in [mm] | Average displacement in the range [mm] |
|---|---|---|
| 20 | 0-5 | 4.0 ± 1.0 |
| 20 | 5–7.5 | 6.7 ± 0.7 |
| 20 | 7.5-10 | 9.0 ± 0.6 |
| 20 | 10-15 | 12.8 ± 1.9 |
| 20 | 15-20 | 17.6 ± 1.4 |

The same set of displacements was used in all the experiments. In each experiment we used each of the displacements as an initial displacement from the gold standard and then applied the registration algorithm. If the resulting TRE, as measured by the residual displacement of the gold-standard points, was smaller than a specified threshold, the registration was pronounced successful. Therefore, the final result is the percentage of successful registrations for each of the displacement ranges. We used a set of thresholds (2 mm, 3 mm, 5 mm, 7 mm and 10 mm) to illustrate how the registration would perform for different clinical treatments, allowing different tolerances. The selection of the thresholds was mainly motivated by the feedback from the radiotherapy experts. The threshold of 3 mm corresponds to the upper boundary of our gold standard ((2 ± 1) mm).

# 3. Results

### *3.1. Texture-feature-based registration performance*

First, we performed intensity-based registration as a part of our experimental work to be able to compare it with our texture-feature-based registration. The success rates (SR%) for intensity-based registration are shown in Table 5.

**Table 5.** Registration success rate (SR%) using intensity features. The values of SR% are given for five different thresholds.

| Range of displacements [mm] | Thresholds [mm]: 2 | 3 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| 0-5 | 7% | 22% | 31% | 33% | 36% |
| 5-7.5 | 5% | 15% | 20% | 22% | 25% |
| 7.5-10 | 5% | 14% | 20% | 23% | 28% |
| 10-15 | 4% | 11% | 15% | 16% | 19% |
| 15-20 | 7% | 10% | 11% | 11% | 12% |

Our proposed method was tested as follows. Using Algorithm 1, we selected two features for two-stage registration in each run. According to the leave-one-out cross-validation principle, selection was made from ten image pairs. Then, the image pair that was not used for the feature selection was registered using the selected features.

Algorithm 1 selected the same robust feature in all the test runs. It is clear from the first row of Fig. 6 that the selected feature is heavily smoothed and depicts the main anatomical structures in both images. A feature with similar characteristics would probably be intuitively selected if the selection had been manual. The choice of accurate feature alternated between two features, one of which is shown in the second row of Fig. 6. Neither of the features is heavily smoothed and both include only detailed structures. This is important for the registration accuracy. Note also that the accurate feature, derived from EPI image, as shown in the second row in Fig. 6, contains a slightly visible grid-like pattern. This pattern is invisible on the original images, but it cannot be avoided, as it is caused by the mechanical setup of the imaging device. However, since there are no corresponding artifacts on the DRR images, it should have no significant influence on the registration. The results for the selected features are shown in Table 6.

**Table 6.** SR%, obtained using two-stage registration with features selected according to Algorithm 1. The values of SR% are given for five different thresholds.

| Range of displacements [mm] | Thresholds [mm]: 2 | 3 | 5 | 7 | 10 |
|---|---|---|---|---|---|
| 0-5 | 37% | 86% | 96% | 100% | 100% |
| 5-7.5 | 35% | 75% | 85% | 91% | 98% |
| 7.5-10 | 28% | 67% | 79% | 86% | 94% |
| 10-15 | 18% | 40% | 49% | 56% | 64% |
| 15-20 | 10% | 18% | 20% | 23% | 25% |

# 4. Discussion

It is evident from Table 5 that intensity-based-registration was not able to properly register images, even for the smallest initial displacements. The intensity-based method reached, at best, a 36% success rate for a very loose threshold of 10 mm. Even for low initial displacements this method diverged significantly – for an initial displacement range of 0–5

mm almost 70% of the results fell outside the 7-mm threshold. This test illustrates that the combination of DRR and EPI images is not a trivial registration problem.

It is clear that the registration using the proposed method (Table 6) performs significantly better than intensity-based registration. For example, the registration results for initial displacements of 5–7.5 mm were, for the proposed method, successful in 75% of cases for the clinically motivated threshold of 3 mm, whereas the intensity-based registration succeeded in only 15% of cases. Even for the most demanding threshold of 2 mm it performs better than the intensity-based registration, regardless of the threshold used.

In addition to the above-described experiments, individual elements of the proposed method were tested as well. Those results are briefly summarized to keep the presentation concise.

In all cases, discarding the principles outlined in Algorithm 1 yielded significantly worse results than shown above in Table 6. For example, when the worst features (quality-wise, as determined by $Q_{RON}$ and $Q_{ACC}$) were used for the registration instead of those recommended by Algorithm 1, the SR% reached only 15% for a threshold of 3 mm and an initial range of displacements of 5–7.5 mm. This is in stark contrast to the results of our method, where this value reached 75%. From looking at the worst features, as shown in the lower half of Fig. 6, this is not surprising. It is clear that they have very little resemblance to the anatomical structures of the pelvis, especially when compared to the best features, shown in the upper half of Fig. 6.

The two-stage registration with the features that violated the daisy-chain condition from line 22 in Algorithm 1 also performed badly. For a threshold of 3 mm and an initial range of displacements of 5–7.5 mm the SR% reached only 55%.

To illustrate the importance of the multi-stage method, we performed single-stage registration experiments. In the first experiment the robust feature from the two-stage method was used alone, and in the second experiment the accurate feature from the two-stage method was used alone. The proposed two-stage method outperformed both single-stage registrations with either of the best features. In the single-stage registration with the robust feature, the SR% reached 75% for a threshold of 3 mm and initial range of displacements of 5–7.5 mm, which is the same performance as in our method. However, for a threshold of 2 mm, the SR% reached only 19%, which is much worse than with our method, where the comparable value was 35%. A single-stage registration with the accurate feature yielded a SR% of 31% for a threshold of 3 mm and an initial range of displacements of 5–7.5 mm, which is again significantly worse than with our method.

# 5. Conclusions

We have presented a novel method for the multi-stage texture-based registration of medical images. Some images, for example, the DRR and low-contrast EPI images used in this study, proved to be difficult to register using intensity information alone. Thus, texture-based registration was used instead. There is a myriad of texture features available for registration; however, we cannot expect that the same texture feature would be universally appropriate. Our framework provides the means for a consistent selection of the most appropriate features for registration from the large number of texture features. We designed a multi-stage registration method that allows robust registration using robust but inaccurate features in the first stage, and then daisy-chaining successively less robust, but more accurate features.

In this study, rotationally invariant Laws' texture features were used. Nevertheless, other texture features could be used or added without affecting other components of the proposed method. Moreover, it could be viewed as a flexible way to fuse many different

image features. In these experiments we have demonstrated that our approach takes advantage of either robust or accurate features.

In general, the texture-based registration performed better than the intensity-based registration. We also showed that not all the texture features are appropriate for a given registration task. The experiments have shown that, overall, for our image data set, the multi-stage registration provides better results than single-stage registration.

The main drawback of our method is that a sufficiently large data set of representative image pairs with the available gold standard is available. On the other hand, this data set is used in the selection step not only to select texture features for later registration, but also for estimating whether the available set of features is sufficient for the particular class of registration problems. The selection step can already point to individual image pairs that diverge from the rest of the data set. Thus, we can predict the performance of the registration, before the actual registration is performed. Our method also assumes that the image data set available for the selection step is representative. We expect that to obtain a representative data set the procedures of the image acquisition would have to be standardized to a significant degree, to minimize the variations in the obtained images.

In our future work we plan to evaluate the proposed method using a larger data set and a larger number of texture features, which will, consequently, demand more computational power.

## Acknowledgments

The authors would like to thank the Institute of Oncology, Ljubljana, for the provision of the image data set. Particular thanks go to P. Petrič MD, M.Sc. for his encouragement and efficient cooperation. The authors also thank P. Rogelj PhD for kindly supplying them with the image-registration software. Our work was financially supported by the Slovenian Ministry of Higher Education, Science and Technology, under grant 3211-05-000557.

## References

[1] Van den Elsen PA, Maintz BA, Pol EJD, Viergever MA: Automatic registration of CT and MR brain images using correlation of geometrical features. IEEE Trans Med Imaging 14(2): 384-396, 1995

[2] Maintz A, Viergever MA: A survey of medical image registration. Med Image Anal 2(1):1–36, 1998

[3] Lester H, Arridge SR: A survey of hierarchical non-linear medical image registration. Pattern Recogn 32(1):129–149, 1999

[4] Hill DLG, Batchelor PG, Holden M, Hawkes DJ: Medical image registration. Phys Med Biol, 46:1–45, 2001

[5] Pluim JP, Maintz BA, Viergever MA: Mutual information based registration of medical images: a survery. IEEE Trans Med Imaging 22(8):986-1004, 2003

[6] Zitova B, Flusser J: Image registration methods: a survey. Image Vision Comput

21(11):977–1000, 2003

[7] Gholipour A, Kehtarnavaz N, Briggs R, Devous M, Gopinath K: Brain functional localization: A survey of image registration techniques. IEEE Trans Med Imaging 26(4):427–451, 2007

[8] Jarc A, Rogelj P, Kovačič S: Texture feature based image registration. In Symposium proceedings Biomedical imaging:From Nano to Macro, IEEE ISBI, 17–20, 2007

[9] Maurer CR, Aboutanos GB, Dawant BM, Maciunas RJ, Fitzpatrick JM: Registration of 3-D images using weighted geometrical features. IEEE Trans Med Imaging 15(6):836–850, 1996

[10] Likar B, Pernuš F: Automatic extraction of corresponding points for the registration of medical images. Med Phys 26(8):1678–1686, 1999

[11] Pennec X, Ayache N, Thirion JP: Landmark-based registration using features identified through differential geometry. Inc. Orlando, FL, USA: Academic Press, 2000

[12] Viola P, Wells WM: Alignment by maximization of mutual information. Int J Comput Vision 24(2):137–154, 1997

[13] Rousseau F, Fablet R, Barillot C: Robust statistical registration of 3D ultrasound images using texture information. IEEE International Conference on Image Processing (ICIP) 1:581–584, 2003

[14] Pratikakis I, Barillot C, Hellier P: Robust multi-scale non-rigid registration of 3-D ultrasound images. In Int. Conf. on Scale-Space and Morphology in Computer Vision, Springer-Verlag, 389–397, 2001

[15] Jarc A, Rogelj P, Kovačič S: Analysis of texture features for registration of DRR and EPI images. Int J Comput Assisted Radiol Surg, 2(Supplement 1):116–118, 2007

[16] Pluim JP, Maintz BA, Viergever MA: Image registration by maximization of combined mutual information and gradient information. IEEE Trans Med Imaging 19(8):809–814, 2000

[17] Maintz BA, Van den Elsen PA, Viergever MA: Comparison of edge-based and ridge-based registration of CT and MR brain images. Med Image Anal 1(2):151-161, 1996

[18] Hsu LY, Loew MH: Fully automatic 3-D feature-based registration of multi-modality medical images. Image Vision Comput 19(1):75–85, 2001

[19] Liu J, Vemuri BC, Marroquin JL: Local frequency representations for robust multimodal image registration. IEEE Trans Med Imaging 21(5):462-469, 2002

[20] Haber E, Modersitzki J: Intensity gradient based registration and fusion of multi-modal images. In Med Image Comput Comput-Assisted Intervention-MICCAI, Springer-Verlag:726-733, 2006

[21] Liu J, Tian J: Registration of brain MRI/PET images based on adaptive combination of

intensity and gradient mutual information. Int J Biomed Imaging (ID 93479), 2007

[22] Škerl D, Likar B, Pernuš F: A protocol for evaluation of similarity measures for rigid registration. IEEE Trans Med Imaging 25(6):779-791, 2006

[23] Reinstein LE, Amols HI, Biggs PJ, Droege RT, Filimonov AB, Lutz WR, Shalev S: Radiotherapy portal imaging quality. Published for the American Association of Physicists in Medicine by the American Institute of Physics, Report of AAPM task group No. 28, 1987

[24] Škerl D, Tomaževič D, Likar B, Pernuš F: Evaluation of similarity measures for reconstruction-based registration in image-guided radiotherapy and surgery. Int J Radiat Oncol Biol Phys 65(3):943–953, 2006

[25] Škerl D, Likar B, Fitzpatrick JM, Pernuš F. Comparative evaluation of similarity measures for the rigid registration of multi-modal head images. Phys Med Biol 52:5587–5601, 2007

[26] Jarc A, Perš J, Rogelj P, Kovačič S: Efficient sampling for the evaluation protocol for 2-D rigid registration. Informacije MIDEM in press, 2008.

[27] Laws K: Rapid texture identification. SPIE Image Processing for Missile Guidance; Proceedings of the Seminar, San Diego, CA: 376–380 1980

[28] Petrou M, Sevilla PG. Image Processing: Dealing with Texture, Hoboken, NJ: John Willey and Sons, 2006

[29] Press WH, Flannery BP, Teukolsky SA, Vetterling WT: Numerical Recipes in C: The Art of Scientific Computing, second edition, Cambridge University press, reprinted in 1996

[30] Dong L, Boyer A: An image correlation procedure for digitally reconstructed radiographs and electronic portal images. Int J Radiat Oncol Biol Phys 33(5):1053–1060, 1995

[31] Matsopoulos G, Asvestas P, Delibasis K, Kouloulias V, Uzunoglu N, Karaiskos P,. Sandilos P: Registration of electronic portal images for patient set-up verification. Phys Med Biol, 49:3279–3289, 2004

[32] Bachman KA: Mutual information-based registration of digitally reconstructed radiographs and electronic portal images. Master's thesis, Applied Mathematics, University of Colorado at Denver, 2005

[33] Khamene A, Bloch P, Wein W, Svatos M, Sauer F. Automatic registration of portal images and volumetric CT for patient positioning in radiation therapy. Med Image Anal, 10:96–112, 2006

[34] Ureten O, Erkal HS: Measurement of patient setup errors using digitally reconstructed radiographs and electronic portal images. Biomedical Engineering Days, Proceedings of the 1998 2nd International Conference 88-90, 1998