EFFICIENT SAMPLING FOR THE EVALUATION PROTOCOL FOR 2-D RIGID REGISTRATION

Andreja Jarc[1,2], Janez Perš[2], Peter Rogelj[2], Stanislav Kovačič[2]
[1]Sipronika d.o.o., Ljubljana, Slovenia
[2] University of Ljubljana, Faculty of Electrical Engineering, Machine Vision Laboratory

Keywords: image registration, evaluation protocol, pseudo-random vs. quasi-random sampling.

Abstract: In our research we aim to reduce the computation time spent on the evaluation protocol of a criterion function for rigid registration tasks. The basic evaluation protocol is performed on *N* uniformly distributed sampling lines in the *K*-dimensional transformation space. Similarity between two images is measured at each of the equidistantly placed points on each sampling line, which is a computationally intensive process. We hypothesize that the computational complexity can be reduced if attention is paid to the selection method of the sampling lines. In the research we compared four sampling methods which affect density and distribution of sampling lines: basic regular sampling, pseudo-random, Sobol quasi-random and Halton quasi-random sampling. We show that the use of Halton quasi-random generator yielded the most uniform distribution of sampling line. Thus, with proper sampling, the number of sampling lines could be systematically reduced in comparison to pseudo-randomly generated lines. This would result in shorter computation time spent on the protocol. The reduction of computation time is especially important when many criterion functions need to be evaluated (e.g. texture feature based image registration). The evaluation protocol was tested on a set of 11 2-D DRR (Digital Reconstructed Radiograph) and EPI (Electron Portal Image) image pairs. The tests have been conducted on intensity feature images as well as texture feature images extracted from the original intensity images. The paired Student's *t*-tests ($p<0.05$) indicated that results obtained from Halton quasi-random sampling were statistically significantly more consistent than the results based on regular or pseudo-random sampling.

UČINKOVITO VZORČENJE ZA VREDNOTENJE KRITERIJSKIH FUNKCIJ ZA 2-D TOGO PORAVNAVO SLIK

Ključne besede: poravnava slik, vrednotenje kriterijskih funkcij, pseudo-nakjučno vs. kvazi-naključno vzorčenje.

Izvleček: V svojih raziskavah poskušamo zmanjšati čas, ki je potreben za kvantitativno vrednotenje kriterijskih funkcij pri togi poravnavi slik. Osnovni protokol za vrednotenje kriterijskih funkcij temelji na *N* naključno porazdeljenih vzorčnih premicah v *K*-dimenzionalnem prostoru parametričnega transformacijskega modela. Podobnost med slikama izračunamo v ekvidistančnih točkah, ki ležijo na *N* vzorčnih premicah. Računanje podobnosti med slikama je računsko potraten postopek. Predpostavljamo, da bi računsko zahtevnost postopka zmanjšali, če uspemo najti optimalno metodo vzorčenja za generiranje vzorčnih premic. V ta namen med seboj primerjali štiri različne metode vzorčenja, ki vplivajo na gostoto in porazdelitev vzorčnih premic v prostoru. Metode vzorčenja, ki smo jih v testih med seboj primerjali, so bile naslednje: enakomerno vzorčenje po mreži, pseudo-naključno, Sobol kvazi-naključno in Haltonovo kvazi-naključno vzorčenje. Iz rezultatov testov se je izkazalo, da nam Haltonovo kvazi-naključno vzorčenje omogoča najbolj enakomerno in s tem »pravično« porazdelitev vzorčnih premic v prostoru. Enakomerna porazdelitev premic bi nam omogočila, da bi število vzorčnih premic lahko sistematično zmanjšali v primerjavi s številom pseudo-naključno generiranih premic. Zmanjšanje števila premic bi direktno pomenilo skrajšanje računskega časa, potrebnega za vrednotenje kriterijskih funkcij. Pohitritev postopka za vrednotenje kriterijskih funkcij še posebej pride do izraza v primerih, kjer je potrebno ovrednotiti večje število kriterijskih funkcij, za primer, poravnava s teksturnimi značilnicami. Teste različnih vzorčenj smo izvedli na setu 11-ih dvodimenzionalnih DRR (Digital Reconstructed Radiograph) in EPI (Electron Portal Imaging) slikovnih parih. Teste smo izvedli tako na originalnih svetlostnih slikah kot tudi na slikah teksturnih značilnic. Studentov t-test ($p<0.05$) je pokazal, da so rezultati, dobljeni na podlagi Haltonovega kvazi-naključnega vzorčenja statistično signifikantno bolj konsistentni od rezultatov, dobljenih z uporabo pseudo-naključnega vzorčenja.

1 Introduction

Clinical diagnosis, as well as therapy planning and evaluation rely increasingly on multiple images of different modalities. For example, in radiation therapy planning a CT (computer tomography) scan is needed for dose distribution calculations, while the contours of the target lesion are often best outlined on MRI (magnetic resonance image) /1/. Image registration is a procedure, where images of the same anatomical structures, acquired using the same or different imaging devices, are brought into the best possible spatial correspondence with respect to each other. Image registration is therefore a fundamental step of information integration. 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 task of finding such transformation parameters that maximize or minimize a criterion function (CF), which measures a similarity between images as a function of registration. A criterion function can be considered as a function mapping from *K*-dimensional continuous space to a subset of a real line, where K is the number of parameters (degrees of freedom) of the parametrical spatial transformation model /8/. For example, for rigid registration of two-dimensional (2-D) or three-dimensional (3-D) images the value of K is 3 or 6, yielding in 3-D or 6-D optimization problem, respectively. The outcome of registration heavily depends on the criterion function profile.

The quality of CF in terms to registration, as proposed by Škerl et al, can be described by the following parameters: accuracy (*ACC*), risk of non-convergence (*RON*), distinctiveness of the global extremum (*DO*) and capture range (*CR*) /8/. First, the accuracy of the CF is defined as a distance from an optimum of the CF to the 'gold standard' transformation, which corresponds to the

aligned position of images. Next, the risk of non-convergence is a measure of robustness of CF. It includes the number, position and distinctiveness of local extrema. RON also describes how sensitive is a CF to interpolation, sampling, partial image overlap and noise. A *DO* measures how distinctive is a maximum (minimum) in respect to the decreasing (increasing) values of CF away from the optimum. Capture range is referred to a limited range of transformations around the optimum for which CF is a monotonic decreasing (increasing) function.

Exhaustive search of the parametrical transformation space would be an obvious and the most precise method to evaluate CF prior to registration at every transformation estimate in *K*-dimensional space. However, in terms of computational demands, this approach is prohibitory expensive.

Let us take for example a simple 2-D rigid registration problem with *K*=3 transformation parameters (two translations and one rotation), with a grid step size of 1 mm and a capture range of 50 mm. For these modest requirements we would get 50^3 (=125 000) transformation estimates at which CF should be evaluated. If the same requirements were to be met for a 3-D registration problem with *K*=6 degrees of freedom (3 translations, 3 rotations), $1.56 \cdot 10^{10}$ estimates of CF would follow.

The protocol for evaluation of similarity measures for rigid registration /8/ is an improvement of the exhaustive search method, as it applies random sampling to the parametrical space. The protocol has been tested for various multi-modal rigid registration tasks, therefore it is becoming a reference method for evaluation of similarity measures. It was devised by Škerl et al /8,9,10/. The continuous *K*-dimensional space is first normalized so that equal changes of each of the parameters in the normalized parametrical space produce the same mean voxel shift. Next, the normalized *K*-dimensional space is "pierced" by *N* randomly selected lines, where the intersection points with a hyper-sphere are uniformly distributed on the surface of the hyper-sphere with radius *R*. All sampling lines converge 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 points is defined as (2R/M). Let's denote X_0 as the origin or GS transformation and $X_{n,m}$ as one of the sampled points. Each of $X_{n,m}$ represents a *K*-dimensional vector of transformation parameters (see (Figure 1 [1]))

For the evaluation protocol as described above, the density and distribution of sampling lines are crucial to obtain representative estimates of the continuous *K*-dimensional transformation space. Following the recommendations in /8/, the number of sampling lines *N*, defined by a pseudo-random generator should be set to 50 for a 6-D optimization task. This way, sampling points should be uniformly distributed on the surface of a sphere.

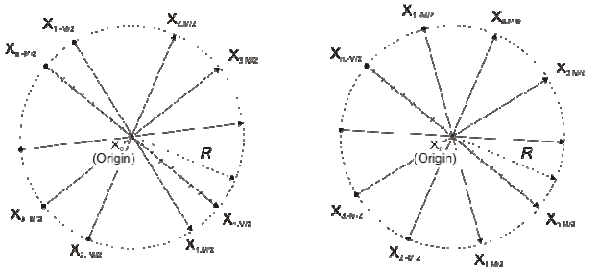


Fig. 1. 2-D parametrical space, sampled by *N* lines and *M* points per line. The maximal displacement from the GS is denoted by *R*, which is a radius of the K-dimensional hyper-sphere. M and R define the step size between sampling points: 2·R/M. The left figure depicts sampling lines, which directions are generated by pseudo-random generator. The right figure depicts sampling lines, which are maximally avoiding each other.

Nevertheless, examples in the literature show that pseudo-random generator is not the optimal choice to fill a *n*-dimensional space uniformly. Sequences of *n*-tuples that fill *n*-space more uniformly than uncorrelated random points are called quasi-random sequences /11/. The main property of sample points given by a quasi-random sequence is that the points are maximally avoiding each other. By this means the same number of sampling lines defined by a quasi-random sequence should cover *K*-dimensional space more evenly than if lines were defined by a pseudo-random sequence.

The following example (Figure 2) shows 2500 points on a sphere, generated by four different sampling methods. The first one is an example of a basic regular sampling of a sphere, the second one is an example of the pseudo-random generated points, the third one for the Sobol quasi-random /12/ and the last one for the Halton quasi-random sampling method /13/.

One can notice, that point density increases towards the poles when the basic regular sampling of a sphere is used. Furthermore, the pseudo-random generated points build small clusters and the vacant space between them is evidently large. The Sobol quasi-random generator delivers more uniform distribution of the points and less vacant space between them. One can see, that the most uniform distribution of the points is generated by Halton quasi-random generator.

The aim of this paper is to evaluate the performance of the evaluation protocol by comparing the consistency of its results for three random sampling methods: pseudo-random, Sobol quasi-random and Halton quasi-random. In addition, the comparison has been performed for the use of basic regular sampling method. Our goal is to find the sampling method that would exhibit similar consistency as the accepted pseudo-random sampling, but with significantly reduced number of sampling points. The comparisons were conducted on a set of 11 2-D DRR (Digital Reconstructed Radiograph) and EPI (Electron Portal image) images. The evaluation protocols have been applied not only to criterion functions based on intensity features but also to several texture feature images extracted from original intensity

[1]The figure is upgraded from /8/.

images/14/.

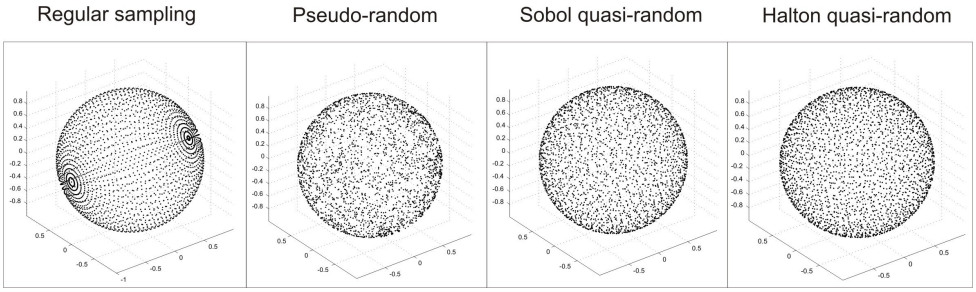


Fig. 2. 2500 points generated by four sampling methods.

Generally, a large number of texture features may be extracted from intensity images. Therefore, the evaluation protocol should be conducted as many times as many features (or potentially their combinations) are available. In general, this may be more than 100 times. If the number of sampling lines was reduced in comparison to the pseudo-random sampling, the time spent on the protocol would be reduced as well and one could evaluate considerably larger number of criterion functions.

We anticipate that the modified evaluation protocol - as proposed in this paper - will be used for the evaluation of a larger number of texture feature based criterion functions prior to registration. The final goal is to select the most appropriate features for a specific registration task. Reduced computational time directly increases the number of texture features that we can afford to evaluate. This increases the chance the best features will be found.

This paper is organized as follows: first, the generation of sampling lines is described in detail. Then, the design of experiments is presented and the data set on which the tests have been conducted is introduced. Finally, some details about texture features used for registration are explained, and the comparisons of results among different sampling methods are shown. Discussion and conclusions complete the paper.

2 Methods and materials

2.1 Generation of sampling lines

In our modified evaluation protocol, the sampling lines in 3-D parametrical space (two translations and one rotation) are generated first by use of the Sobol quasi-random and second by Halton quasi-random generator. To compare with random sampling methods, the 3-D parametrical space is additionally sampled by a basic regular sampling.

The azimuthal angle φ and the polar angle θ of spherical coordinates are the outputs of the regular sampling or the quasi-random generators. r is a distance (radius) from a point to the origin (Figure 3).

The spherical coordinates (r, φ, θ) are related to the Cartesian coordinates (x, y, z) by the following equations /15/:

$$x = r\cos(\varphi)\sin(\theta) \quad (1)$$

$$y = r\sin(\varphi)\sin(\theta) \quad (2)$$

$$z = r\cos(\theta) \quad (3)$$

where $r \in [0,\infty)$, $\varphi \in [0,2\pi]$ and $\theta \in [0,\pi]$.

Each of the N sampling lines in 3-D parametrical space is defined by randomly selected starting position $X_{n,-M/2}$ on a sphere at the distance R from the origin and its mirror point $X_{n,M/2}$. The starting points are defined by a 3-D vector $[x,y,z]$ (Eq. (1), (2) and (3)).

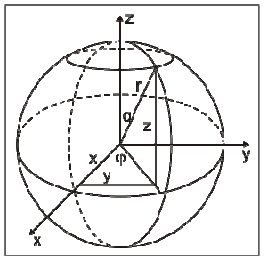


Fig. 3. Our notation of 3-D spherical coordinates.

The following two examples (Figure 4) show 2500 points generated by the method as described above. In the first example, the Cartesian coordinates of the points are defined by the Sobol quasi-random generator and in the second example by the Halton quasi-random generator.

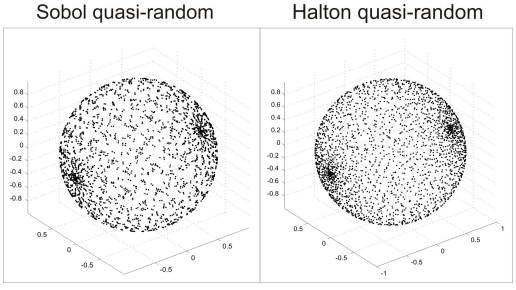


Fig. 4. Distribution of sampling points, where angles φ and θ have been selected by quasi-random generators.

It can be easily seen that points are not uniformly distributed over the sphere surface since they

group in two clusters at the poles. The reason is, that the area element $d\Omega = sin(\theta)d\theta d\varphi$ depends on θ, and therefore points selected in this way are clustered near the poles /16/. To obtain points such that any small area on the sphere is expected to contain the same number of points, we choose u and v to be quasi-random variants on [0,1]. Then:

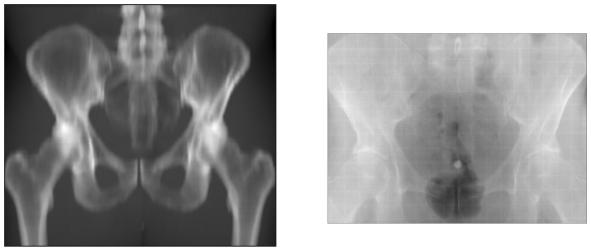
$$\varphi = 2\pi u$$
$$\theta = arccos(2v - 1) \quad (4)$$

gives the spherical coordinates for a set of points, which are uniformly distributed over Ω.

Using this correction, the uniformity of sampling points improves as shown in Figure 2. The above correction was used throughout the paper for all but regular sampling methods.

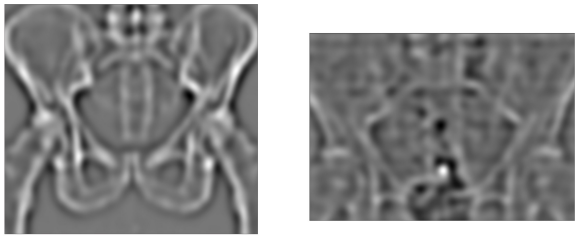
2.2 Test image set

The tests have been conducted on 11 pairs of DRR (Digital Reconstructed Radiograph) and EPI (Electron Portal Imaging) images of the pelvis (Figure 5). By correctly matching the two modalities, it is possible to verify the positioning of the patient during radiation therapy and automatically adjust the positioning if necessary.



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

Fig. 5. An example of one of the intensity image pair. (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.



(a) Laws texture features from DRR (b) Laws texture features from EPI

Fig. 6. (a) Laws texture feature extracted from the intensity image of the reference DRR image. 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. (b) Laws texture feature extracted from the intensity image of the floating EPI image.

The registration of DRR/EPI images is not a trivial problem due to 2-D representation of 3-D data. Several papers have been published proposing and/or investigating various registration methods using these image modalities for patient positioning applications /17,18,19/. However, we found that intensity based registration is not reliable, since the intensity features do not comply with some global intensity relationship, expected by intensity-based registration approaches/20/. Therefore, an alternative registration approach based on texture features has been proposed to register DRR/EPI images/14/.
The images used for the tests were initially aligned as they were used during the radiotherapy practice. The image alignment is achieved with employment of three lasers (sagittal, coronal and axial) for marking the patient reference coordinates/21/. The gold standard-GS registration in our tests was a transformation vector 0 with its tolerance of 3 mm. However, due to the design of our experiments imprecise GS registration did not influence our results.

2.3 Experiment Design

The comparison between the reference evaluation protocol, as implemented by the authors /22/ and three modified versions of the protocol has been performed. The modified versions used regular sampling, Halton and Sobol quasi-random sampling instead of pseudo-random sampling.

First, tests on the intensity images have been conducted. We assume, that images of each image pair are initially aligned. Mutual information (MI) /23,24,5/ has been used to measure similarity between the reference and the transformed floating image at each sampling estimate $X_{n,m}$. MI was estimated from joint density, which was approximated by using a Parzen kernel. The Parzen kernel was applied to the 2-D joint histograms, quantized into 256 x 256 bins. Each joint histogram was created by plotting an *(i,j)* point for every pair of corresponding pixels in the overlap region, where i was the pixel intensity in the reference image, and j was the interpolated pixel intensity of the floating image.

The experiments were designed exactly as described in /22/ except for the number of sampling lines N and the way those lines were generated. The number of sampling lines has been systematically lowered from the recommended value of 50 in steps of 5: N=50, 45, 40, ..., 10, 5, 1. The normalization parameters used to generate sampling lines are listed in Table 1. For the parameter explanation see /8/.

Each sampling line provides a transformation profile of a criterion function. To observe the behavior of criterion functions we checked two parameters: accuracy (*ACC*) and risk of non-convergence (*RON*). Accuracy as it is defined in /8/ is a root mean-square of distances between the maximum value of a criterion function $X_{n,max}$ and origin X_0 on each of the n sampling lines; $n=1,2,\ldots N$.

$$ACC = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left\|X_{n,\max} - X_0\right\|^2} \quad (5)$$

Tab.1. Floating image sizes, pixel sizes, translation and rotation units of normalized parametrical space, radius R, number of points along a line M and a step size between two sampling points.

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

Risk of non-convergence $RON(r)$ is defined as the average of positive gradients $d_{n,m}$ within distance r from each of the N global maxima:

$$RON(r) = \frac{1}{2rN}\sum_{n=1}^{N}\sum_{m=\max-k}^{\max+k} d_{n,m} \quad (6)$$

In our tests, the consistency of *ACC* and *RON* values between sampling lines has been compared for four different sampling methods. Furthermore, the consistency check of *ACC* and *RON* values have been performed for reduced numbers of sampling lines.

Moreover, the original and the modified evaluation protocols have been applied to the texture feature images (Figure 6) and the results have been compared. The conveyed texture information between texture images was measured again by MI.

The experiment details are identical as described above. Again, the effect of using the pseudo-random, regular or quasi-random generator was observed, while lowering the number of sampling lines from 50 in steps of 5.

2.4 Texture features used for registration

Apart from intensity images, the evaluation protocols have been tested on Laws texture features, which were extracted from both of the original intensity images. Laws /25/ developed a set of two-dimensional filter masks, which are composed of combinations of several one-dimensional filters /26/.

For the tests we chose Laws texture features extracted by combinations of level-L and spot-S filter masks. Both 1-D filters were of size 20 mm. L-S texture features were summed up with texture features obtained by S-L filter masks /25/, again of size 20 mm.

Each filtered image was subsequently converted to a texture energy image. Texture energy image was obtained by convolving the local texture feature image by a Gaussian averaging window. We used a Gaussian kernel of size 20 mm and cut off frequency at 3σ. Finally, the 2% extreme values of each texture energy image were saturated and the rest were scaled from 0 to 255 integer level yielding 8-bit quantization. See Figure 6.

2.5 Criterion functions

Our criterion functions are computed by measuring the conveyed intensity and texture feature information between images being registered. We choose mutual information (MI) to measure similarity between images. MI can be computed by using the following formula:

$$MI(A,B) = H(A) + H(B) - H(A,B) \quad (7)$$

with $H(A)$ and $H(B)$ are the Shannon entropies of image features for both of the images and $H(A,B)$ is their joint entropy. Entropy $H(\cdot)$is computed as:

$$H(\cdot) = -\sum_i p(i)\cdot log_2 p(i) \quad (8)$$

where p is a probability distribution of features on an image.

3 Results and discussion

3.1 Test on intensity features

The tests are performed on 11 DDR/EPI image pairs (Figure 7, 8) following this protocol:

1. For each of the 11 DRR/EPI image pairs the sampling lines in 3-D transformation space are generated:

- 400 sampling lines obtained by regular sampling of φ and θ,
- 400 sampling lines for the reference evaluation protocol are obtained through the web-interface /22/,
- 400 sampling lines generated by Sobol quasi-random generator, and
- 400 sampling lines generated by Halton quasi-random generator.

2. The criterion function is evaluated in each of the 400 x 400 sampling points $X_{n,m}$.
3. The evaluation parameters of *ACC* and *RON* are - to consider statistics - calculated for the following consecutive sub-ranges of 400 sampling lines:
    - 8 sub-ranges of 50 sampling lines,
    - 8 sub-ranges of 45 sampling lines,
    - 10 sub-ranges of 40 sampling lines,
    - 11 sub-ranges of 35 sampling lines,
    - 13 sub-ranges of 30 sampling lines,
    - 16 sub-ranges of 25 sampling lines,
    - 20 sub-ranges of 20 sampling lines,
    - 26 sub-ranges of 15 sampling lines,
    - 40 sub-ranges of 10 sampling lines, and
    - 80 sub-ranges of 5 sampling lines.

The results are depicted as scatter of *ACC* and *RON* values. The scatter is computed as normalized standard deviation of the consecutive subranges of sampling lines. We expect that better random generator would yield lower scatter of the results. Lower scatter means more consistent results which is the aim of our tests. For purposes of clarity, results are shown only for *N*=50,40,30,20 and 10.

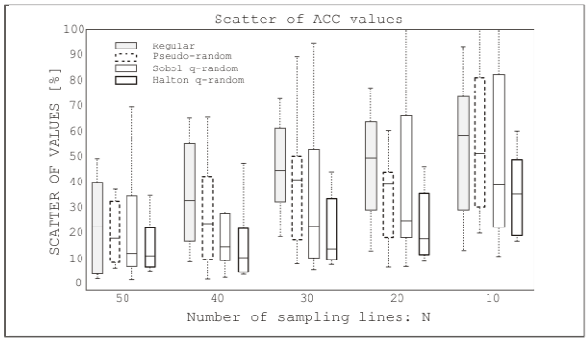


Fig.7. Results for intensity features. Box-and-whisker plots are showing scatter of values of 11 DRR/EPI image pairs for the parameter *ACC* and four sampling methods: regular, pseudo-random, Sobol quasi-random and Halton quasi-random, respectively.

The paired Student's t-test (p<0.05) which compares the scatter of *ACC* values in Figure 7 of the four sampling methods indicated no significant difference between pseudo-random and Sobol-quasi random sampling at any of different numbers of sampling lines. On the other hand, there exists significant difference between results of pseudo-random and Halton-quasi random sampling when the analysis is performed on *N*=40,30,20 and 10 sampling lines. In these cases the pseudo-random generator based results show significant higher scatter of the values in comparison to Halton quasi-random sampling based results. However, there exists no significant difference between results of the two generators when the analysis is performed on *N*=50 and *N*=5 sampling lines. 50 sampling lines seems to be enough in 3-D transformation space to overcome a deficiency of pseudo-random generator, but 5 sampling lines are too few to achieve satisfactory consistence even with Halton quasi-random sampling.

The comparison between regular and pseudo-random sampling shows significantly higher scatter of *ACC* values for *N*=20 when the regular sampling method is employed.

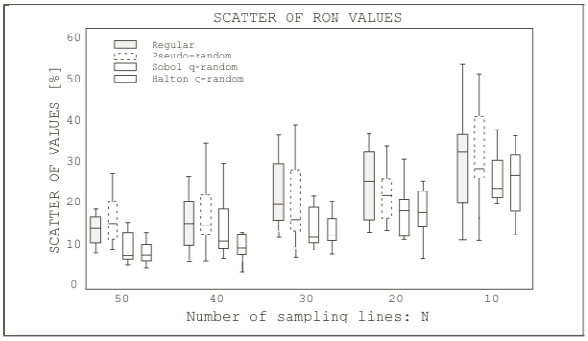


Fig.8. Results for intensity features. Box-and-whisker plots are showing scatter of values of 11 DRR/EPI image pairs for the parameter *RON* and four sampling methods: regular, pseudo-random, Sobol quasi-random and Halton quasi-random, respectively.

Similar to the scatter of *ACC* values, the paired Student's t-test (p<0.05) indicated no significant difference between pseudo-random and Sobol-quasi random sampling based scatter of *RON* values in Figure 8. However, Halton quasi-random sampling yielded significantly lower scatter of *RON* values for all but *N*=5 sampling lines in comparison to pseudo-random sampling. Again, the reason is that 5 sampling lines are too few to cover 3-D transformation space dense enough with any of the generators.

The regular sampling shows no significant difference in the scatter of *RON* values compared to pseudo-random based results for any of *N* sampling lines.

3.2 Tests on texture energy images

The tests are performed on 11 texture image pairs derived from original intensity images (Figure 9, 10). The experiment protocol is the same as the one used for intensity features.

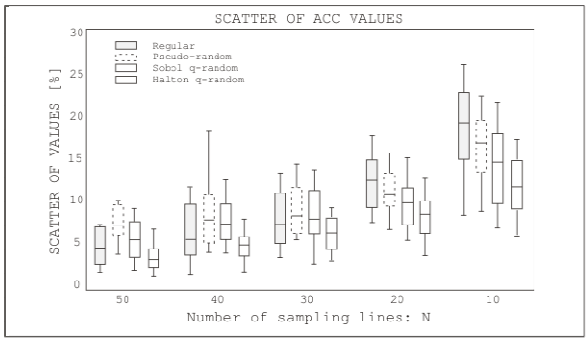


Fig.9. . Results for texture features. Box-and-whisker plots are showing scatter of values of 11 DRR/EPI image pairs for the parameter *ACC* and four sampling methods: regular, pseudo-random, Sobol quasi-random and Halton quasi-random, respectively.

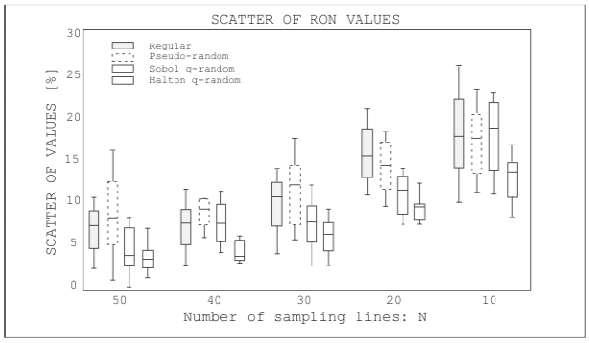


Fig.10. Results for texture features. Box-and-whisker plots are showing scatter of values of 11 DRR/EPI image pairs for the parameter *RON* and four sampling methods: regular, pseudo-random, Sobol quasi-random and Halton quasi-random, respectively.

Note, that overall, the scatter values are much lower for texture features in comparison to intensity features. This is a strong argument to use texture features instead of intensities for this registration task. However, from the tests performed on the texture feature images similar conclusions may be drawn than from the results for intensity features. Again, the Halton quasi-random sampling outperformed both regular and pseudo-random sampling as it yielded more consistent results for both, *ACC* and *RON* values. Student t-test indicates significant difference between samplings for all systematically reduced number of sampling lines, except N=30 - *ACC* values and N=5 - *RON* values.

The recommended number of sampling lines N which would yield enough consistent results for our 3-D optimization task was initially not provided. Since our dimensionality was significantly lower than in the original work by Škerl et al, we started our tests with N=50 as reasonably safe margin. From the results of our tests we can hypothesize that by using Halton quasi-random generator the smallest number of sampling lines N, which would yield enough consistent results for 3-D optimization task, is as low as 10. This hypothesis is confirmed by paired Student t-test (p<0.05) which compared scatter results between pseudo-random N=50 sampling lines and systematically reduced N from 50 to 5 for Halton quasi-random generator. t-test indicated that for N=50 and N=40 Halton quasi-random delivered significantly lower scatter values in comparison to pseudo-random sampling with N=50. From N=35 to N=10 Halton quasi-random sampling indicated no significant difference to pseudo-random sampling with N=50. For N=5 Halton quasi-random sampling delivered significantly higher scatters in comparison to pseudo-random with N=50.

For our registration task the recommended pseudo-random sampling with 50 sampling lines yields comparable scatter of results to the Halton-random sampling with 10 sampling lines.

4 Conclusion

The evaluation protocol described in /8/ assesses the quality of similarity measure used in a specific registration problem prior to registration. This is done by evaluating the behaviour of a similarity measure for simulated transformations. The evaluation of a similarity measure includes the following parameters: accuracy, robustness and capture range.

We are using this evaluation protocol for assessment of criterion functions based on the intensity and a bank of texture features – such use requires the evaluation of many criterion functions. It is therefore important that the evaluation protocol is as efficient as possible, while retaining the truthfulness of the results.

The obtained results from our tests show that Halton quasi-random outperformed regular, pseudo-random and Sobol quasi-random sampling for our specific registration task. Additionally, the tests have shown, that if the computational time is of prime concern, the number of sampling lines may be reduced, which yields a significant reduction of computation time spent on evaluation of one CF. Thus, a larger number of CFs may be evaluated to select those, that would provide the best registration.

Additionally, we can also see that values *ACC* and *RON* based on texture features deliver considerably lower scatters than values based on intensity features. It may be concluded that a lower scatter of results may be one of the additional parameters to assess the quality of a criterion function. The lower the scatter of the results the more representative are the criterion functions defined on each of the sampling lines. Moreover, it is less likely that a criterion function is ill-defined by containing a false global optimum and strong local maxima /27/.

5 Acknowledgement

The authors would like to thank the Institute of Oncology Ljubljana for the provision of the image data set and to the Slovenian Ministry of Higher Education, Science and Technology for the financial support, under grant 3211-05-000557.

6 References

/1/ P. A. Van den Elsen, J. B. A. Maintz, E. J. D. Pol, and M. A. Viergever. Automatic registration of ct and mr brain images using correlation of geometrical features. *IEEE Trans. Med. Imag*, 14(2), 1995.

/2/ A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

/3/ H. Lester and S. R. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32:129–149, 1999.

/4/ D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Phys.Med.Biol.*, 46:1–45, 2001.

/5/ W. Pluim, J. P., A. Maintz, and M. A. Viergever. Mutual information based registration of medical images: a survery. *IEEE Trans. Med. Imag.*, 2003.

/6/ B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, 2003.

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

/8/ D. Škerl, B. Likar, and F. Pernuš. A protocol for evaluation of similarity measures for rigid registration. *IEEE Trans. Med. Imag.*, 25(6), 2006.

/9/ D. Škerl, D. Tomaževič, B. Likar, and F. Pernuš. Evaluation of similarity measures for reconstruction-based registration in image-guided radiotherapy and surgery. *Int. J. Radiation Oncology Biol. Phys.*, 65(3):943–953, 2006.

/10/ D. Škerl, B. Likar, J. M. Fitzpatrick, and F. Pernuš. Comparative evaluation of similarity measures for the rigid registration of multi-modal head images. *Phys. Med. Biol.*, (52):5587–5601, 2007.

/11/ H. W. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University press, reprinted in 1996.

/12/ I. M. Sobol. The distribution of points in a cube and the approximate evaluation of integrals. 7(4):86–112, 1967.

/13/ J. H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2(1):84–90, 1960.

/14/ A. Jarc, P. Rogelj, and S. Kovačič. Texture feature based image registration. In *Symposium proceedings*, pages 17–20, Arlington Virginia, U.S.A., April 2007. IEEE ISBI.

/15/ E. W. Weisstein. Spherical coordinates, from mathworld–a wolfram web resource. *http://mathworld.wolfram.com/SphericalCoordinates.html*, accessed on 24.10.2007.

/16/ E. W. Weisstein. Sphere point picking, from mathworld–a wolfram web resource. *http://mathworld.wolfram.com/SpherePointPicking.html*, accessed on 24.10.2007.

/17/ G. Matsopoulos, P. Asvestas, K. Delibasis, V. Kouloulias, N. Uzunoglu, P. Karaiskos, and P. Sandilos. Registration of electronic portal images for patient set-up verification. *Phys. Med. Biol.*, 49:3279–3289, 2004.

/18/ L. Dong and A. Boyer. An image correlation procedure for digitally reconstructed radiographs and electronic portal images. *Int. J. Radiation Oncology Biol. Phys.*, 33(5):1053–1060, 1995.

/19/ A. Khamene, P. Bloch, W. Wein, M. Svatos, and F. Sauer. Automatic registration of portal images and volumetric ct for patient positioning in radiation therapy. *Med. Image Anal.*, 10:96–112, 2006.

/20/ A. Jarc, P. Rogelj, and S. Kovačič. Analysis of texture features for registration of drr and epi images. *International Journal of Computer Assisted Radiology and Surgery*, 2(Supplement 1):116–118, 2007.

/21/ O. Ureten and Erkal H. S. Measurement of patient setup errors using digitally reconstructed radiographs and electronic portal images. pages 88–90. 2nd International Biomedical Engineering Days, 1998.

/22/ D. Škerl, B. Likar, and F. Pernuš. An online protocol for evaluation of similarity measures. *www.lit.fe.uni-lj.si/Evaluation*, 2006.

/23/ F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imag.*, 16(2), 1997.

/24/ P. Viola and W. M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

/25/ K. Laws. Rapid texture identification. pages 376–380. SPIE Image Processing for Missle Guidance, 1980.

/26/ M. Petrou and P. G. Sevilla. *Image Processing Dealing with Texture*. John Willey and Sons, 2006.

/27/ J. Pluim, Maintz J. B. A., and M. Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE Trans. Med. Imag.*, 19(8), 2000.

Andreja Jarc, M.Sc.
Sipronika d.o.o.
Tržaška 2, 1000 Ljubljana
andreja.jarc@sipronika.si

Dr. Janez Perš,
University of Ljubljana
Faculty of Electrical Engineering,
Tržaška 25, 1000 Ljubljana, Slovenia

Dr. Peter Rogelj,
University of Ljubljana
Faculty of Electrical Engineering,
Tržaška 25, 1000 Ljubljana, Slovenia

Prof. Dr. Stanislav Kovačič,
University of Ljubljana
Faculty of Electrical Engineering,
Tržaška 25, 1000 Ljubljana, Slovenia