Regional Image Similarity Criteria Based on the Kozachenko-Leonenko Entropy Estimator

Juan D. García-Arteaga Jan Kybic
Center for Machine Perception, Czech Technical University, Prague, Czech Republic
{garcia,kybic}@fel.cvut.cz

Abstract

Mutual Information is one of the most widespread similarity criteria for multi-modal image registration but is limited to low dimensional feature spaces when calculated using histogram and kernel based entropy estimators. In the present article we propose to use the Kozachenko-Leonenko entropy estimator (KLE) to calculate higher order regional mutual information using local features. The use of local information overcomes the two most prominent problems of nearest neighbor based entropy estimation in image registration: the presence of strong interpolation artifacts and noise. The performance of the proposed criterion is compared to standard MI on data with a known ground truth using a protocol for the evaluation of image registration similarity measures. Finally, we show how the use of the KLE with local features improves the robustness and accuracy of the registration of color colposcopy images.

1. Introduction

Mutual information (MI) was invented by Shannon [21] as a measure of the statistical independence between two signals without assuming any *a priori* relationship between them and re-introduced as a measure of image alignment by Viola and Wells [25] and Collignon *et al.* [4] in the mid-nineties. Over the years, MI has proven a particularly well suited and attractive option for medical image registration as evidenced by its extensive use and *de facto* status as the standard for multimodal medical image registration (see the surveys on MI based medical image registration by Pluim *et al.* [16] and Maes *et al.* [13] for more details).

MI based image registration does fail, however, when images share only a small amount of information or are of poor quality. In this situations the errors may be due to a global minimum located at an incorrect location [15] or to the presence of spurious local minima that trap the optimization algorithm away from the correct solution [3]. As MI is usually calculated using only the intensities of two scalar grayscale images, the use of additional information to improve the robustness and precision of the similarity criterion when registering difficult image pairs has been proposed as a possible solution. A common approach to do this is to include, along with the intensity, additional channels in the MI calculation such as gradients [15][5][23], first and second order texture features [2], labels [22], or coded geometric information [20][12]. The main reported advantages of using these methods [23] are increased robustness and insensitivity to intensity inhomogeneities. As most of these systems use histograms to estimate the MI ([20] and [12] being notable exceptions) the number of features that can be used simultaneously before the histogram becomes too sparse is low.

In the present article we propose the use of high dimensional features together with the Kozachenko-Leonenko entropy estimator [7][8] (KLE) to achieve improved registration robustness. The KLE calculation is based on nearest neighbor (NN) graphs to avoid the lack of precision and high costs of using kernel or histogram estimators in high dimensions. We show how the use of higher order features avoids the main problems of image similarity criteria calculated using NN entropy estimators from finite accuracy data: the presence of strong artifacts and the noisiness of the functions.

The article is organized as follows: Section 1 reviews the concepts of MI, some of its adaptations to regional features and its calculation using the KLE estimator. Section 2 describes the proposed similarity criteria using local features together with the KLE. In Section 3 we quantitatively compare the performance of image similarity criteria calculated using KL against standard sampled MI using a protocol for the evaluation of image registration similarity measures and simulated MRI images with strong inhomogeneities. In Section 4 we present an application of the developed criteria in the registration of colposcopy images. We also show how it is possible to register these images using the full color information using the KLE estimator.

1.1. Mutual Information

The MI criterion J_{MI} from two images f and g is defined in terms of entropy H as:

$$MI(F,G) = H(F) + H(G) - H(F,G) \quad (1)$$

$$J_{MI}(f,g) = -MI(F,G) \quad (2)$$

$$H(F) = -\sum_{\mathbf{x}_i \in \Omega} p_F(f(\mathbf{x}_i)) \log\left(p_F\left(f(\mathbf{x}_i)\right)\right) \quad (3)$$

where F and G are random variables and pixel values at $f(\mathbf{x}_i)$, $g(\mathbf{x}_i)$ at position $\mathbf{x}_i$ are assumed to be independent realizations of F, G :$f(\mathbf{x}_i) \sim F$, $g(\mathbf{x}_i) \sim G$, $(f(\mathbf{x}_i), g(\mathbf{x}_i)) \sim (F,G)$. J_{MI} measures the degree of independence between F and G. Probability distribution functions p_F, p_G, and $p_{(F,G)}$ are, in general, unknown and must be estimated retrospectively.

If f and g correspond to misaligned images of the same scene it is assumed that the minimum of J_{MI} will occur for the transformation $\mathbf{T}$ that aligns them.

$$\mathbf{T}^* = \underset{\mathbf{T}}{\operatorname{argmin}}\, J_{MI}\big(f\left(\mathbf{x}\right), g\left(\mathbf{T}\left(\mathbf{x}\right)\right)\big) \quad (4)$$

Image registration is then defined as the process of finding the transformation $\mathbf{T}$ that minimizes the similarity criterion and removes the geometric distortion.

1.2. High Order MI

Rueckert *et al.*[17] introduced the concept of high-order MI for the non-rigid registration of MR images with strong RF inhomogeneities. The basic idea is to, instead of only using the pixels' intensities, to calculate the MI of two images f, g based on the entropy of the co occurrence of neighboring pixel values. Let $P = \{(\mathbf{x}_i, \mathbf{x}_j)\}$ be the set of all neighboring pixel pairs in image f based on a 4-neighborhood. The second order entropy is then defined as:

$$H_2(F) = -\sum_{(i,j)\in P} p(f(\mathbf{x}_i), f(\mathbf{x}_j)) \log\left(p(f(\mathbf{x}_i), f(\mathbf{x}_j))\right), \quad (5)$$

where $p(fx(\mathbf{x}_i), f(\mathbf{x}_j))$ is the joint probability of the value pair $[f(\mathbf{x}_i), f(\mathbf{x}_j)]$ occurring in neighboring pixels. To calculate $H_2(F)$, a 2D histogram is populated by drawing votes from every pixel and each of it's spatial neighbors (4 for pixels and 6 for voxels). In this histogram votes from homogenous region pixels ($f(\mathbf{x}_i) \approx f(\mathbf{x}_j)$) fall in the bins located in or near the histogram's diagonal. Votes of pixels from regions with sharp intensity changes, *e.g.* boundaries between regions, fall above or below the histogram's diagonal depending on the coordinate choice and the type of intensity change (dark-to-light or light-to-dark). This histogram contains then not only information of the intensity distribution but also partial information of the image structure.

Second order MI is defined as:

$$I_2(F,G) = H_2(F) + H_2(G) - H_2(F,G). \quad (6)$$

The third term is calculated using a 4D histogram estimation of:

$$p_{FG2}(f(\mathbf{x}_i), f(\mathbf{x}_j), g(\mathbf{x}_i), g(\mathbf{x}_j)), \quad (7)$$

where p_{FG2} measures the probability of the joint occurrence of values $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$ in neighboring pixels in image f and of values $g(\mathbf{x}_i)$ and $g(\mathbf{x}_j)$ in the same pixels in image g.

The main drawback of this approach is its need to approximate high-dimensional probability distributions through histograms. As more dimensions are added to the histogram the number of bins grows exponentially while the number of samples remains constant. The histogram then begins to be too sparsely populated and the estimate of the probability becomes less reliable due to the lack of sufficient samples per bin. The authors use 16 bins per dimension as a compromise between the total number of bins and the average number of samples per bin.

Russakoff *et al.* [18] extended the second-order MI definition into a regional mutual information (RMI) that includes all pixels within a radius. To avoid calculating high dimensionality histograms they simplify the problem by assuming that the high dimensional distribution is normal and can be transformed into a space where each dimension is independent. Using these two assumptions they can calculate the entropy as that of a set of normally distributed points with covariance matrix Σ_d:

$$H_g(\Sigma_d) = \log((2\pi e)^{\frac{d}{2}} \det\left(\Sigma_d\right)^{\frac{1}{2}}). \quad (8)$$

1.3. Kozachenko-Leonenko Estimator

The Kozachenko-Leonenko estimator (KLE) [7][8] calculates the entropy from N_S samples of a distribution from the L_2 distance between each sample $\mathbf{f}(n) \in \mathbb{R}^D$ and its nearest neighbor $\mathbf{f}_{NN}(n)$ among the $N_S - 1$ remaining samples:

$$\mathcal{H}_{KL}^{N_S}(\mathbf{f}) = \frac{D}{2N_S} \sum_{n=1}^{N_S} \log(\|\mathbf{f}(n) - \mathbf{f}_{NN}(n)\|) + K, \quad (9)$$

$$\mathbf{f}_{NN}(n) = \arg\min_{x \neq n} \|\mathbf{f}(n) - \mathbf{f}(x)\| \quad (10)$$

with:

$$K = \log \frac{(N_S - 1)\pi^{D/2}}{\Gamma(1 + \frac{D}{2})} + \gamma \quad (11)$$

where γ is the Euler constant and D is the number of dimensions of $\mathbf{f}$.

First proposed for image registration by Kybic [9], the KLE estimator presents several problems that need to be solved before it can be used on a practical image registration application. As seen in (9) the KLE estimator will diverge when a sample's nearest neighbor is itself, *i.e.* when two or more elements have the same value:

$$\mathbf{f_{NN}}(i) = \mathbf{f}(i) \rightarrow \|\mathbf{f_{NN}}(i) - \mathbf{f}(i)\| = 0 \qquad (12)$$

Although for continuous distributions the probability of a "crash" (two samples having the same value) is zero, this is not the case when working with discrete valued images. A modified entropy estimator $\mathcal{H}_{DKL}$ was proposed in [10]:

$$\mathcal{H}^{N_S}_{DKL}(\mathbf{f}) = \frac{D}{2N_S}\sum_{n=1}^{N_S}\log(\max(\|\mathbf{f}(n) - \mathbf{f}_{NN}(n)\|, \epsilon) + K. \qquad (13)$$

The value ϵ is chosen to be below the volume of a quantization bin. Using estimate (13) avoids the divergence problems of (9) but results in an un-smooth MI criterion profile that presents strong interpolation artifacts when the sampling grid points of the two images overlap (Figure 1). The artifacts may be attenuated by filtering off the image's high frequencies thereby minimizing the number of crashes. The resulting function (seen in Figure 1) is, however, still noisy and does not reach its maximum at the correct alignment position.

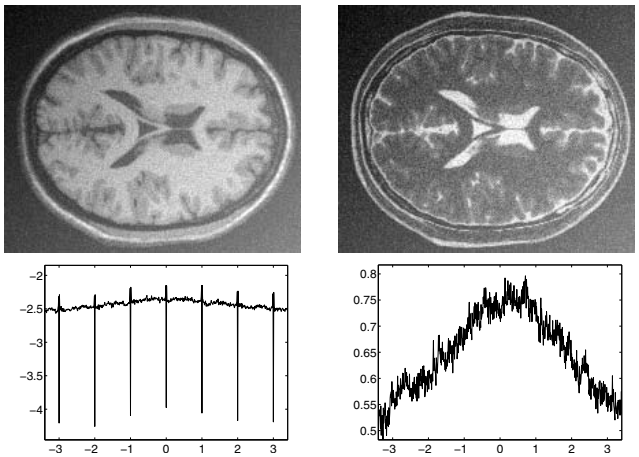

Figure 1. Two axial cuts with a strong bias field from a T2-T1 weighted brain scan simulated with Brainweb [1] (top) are displaced horizontally from their correct alignment to test the MI_{DKL} similarity criterion . The plots of the MI vs. displacement in pixels from correct alignment are calculated directly from the discrete images (bottom left) and from the pre-filtered double-precision images (bottom right) using the DKL entropy estimator.

2. Method

In the present section we retake the basic ideas outlined in Section 1.2 in order to apply the DKL entropy estimator to higher dimensional features. We wish to profit from the estimator's strong convergence, theoretical elegance and ability to manage high-dimensional feature spaces to improve registration results. We show how the noise and artifacts observed in the MI_{DKL} may be removed by choosing an adequate representation of the spatial information, making the resulting function adequate for image registration.

2.1. Data Criteria

To improve registration results we propose to calculate the data information criteria (either entropy or MI) from a feature vector consisting of a pixel and its spatial neighbors using the DKL entropy estimator. For evaluation purposes we will limit ourselves to 2 and 8 spatial neighbors. For two and eight neighbors the evaluated feature vectors of the pixel at location $\mathbf{x}_{ij}$ in a 2D image f will be:

$$\mathbf{f^2}(\mathbf{x}_{ij}) = [f(i,j), f(i+1,j), f(i,j+1)], \qquad (14)$$

$$\mathbf{f^8}(\mathbf{x}_{ij}) = [f(i+d_i, j+d_j)], d_i, d_j \in (-1,0,1). \qquad (15)$$

The proposed criteria to be optimized when registering images f and g using d neighbors will then be the entropy or the MI of vectors $\mathbf{f^d}$ and $\mathbf{g^d}$ calculated using the entropy estimator defined in Equation 13:

$$MI_{DKL} = \mathcal{H}^{N_S}_{DKL}(\mathbf{f^d}) + \mathcal{H}^{N_S}_{DKL}(\mathbf{g^d}) - \mathcal{H}^{N_S}_{DKL}(\mathbf{f^d}, \mathbf{g^d}). \qquad (16)$$

$$H_{DKL} = -\mathcal{H}^{N_S}_{DKL}(\mathbf{f^d}, \mathbf{g^d}). \qquad (17)$$

Notice that the feature vectors $\mathbf{f^d}$, $\mathbf{g^d}$, and $(\mathbf{f^d}, \mathbf{g^d})$, although defined by the matrix indexes, are N_S sets of 3,6, 9 or 18 dimensional vectors, depending on d and wether we are calculating the images' individual or joint entropy, and that the estimated entropy results will be independent of the sample order. This means that the KL formulation may be used directly without any further *ad hoc* adaptations to include spatial information as this information is already *implicitly contained in the feature's intensity values*. This retains the estimator's desirable characteristics such as strong convergence and zero bias as the number of samples grows.

An example showing the differences in the location of the NN found when using only the intensity values and when using 2 neighbors is shown in Figure 2. Several advantages of the latter for calculating the entropy using KL should be noted:

- Abrupt "switches" between nearest neighbors for small deformation changes are fewer and smaller, meaning that the topology of the associated NN graph remains mostly unchanged, making the MI criterion

smoother. A formal demonstration for Minimum Spanning Trees (easily extendable to NN graphs) may be found in [19].

- Entropy estimation becomes a spatially local process due to the increased discriminatory power of the extended image descriptor. This is evidenced by the majority of NN being found very near the original pixel triplet location.

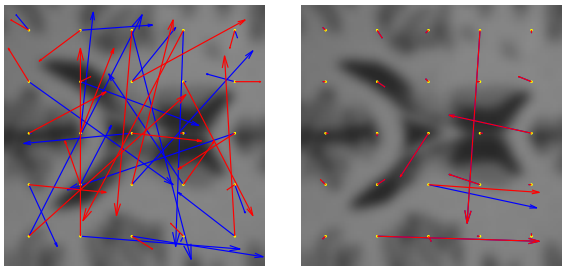

Figure 2. The arrows connect the pixels in the yellow dot grid to the spatial location of their NN in feature space when calculating the joint entropy of a 100x100 window from the images shown in Figure 1. Blue arrows correspond to the NNs found when the T2 image (not shown) is displaced by 2 pixels and red arrows (overlayed) correspond to the NNs found when it is displaced by 1.75 pixels. Results when using the joint intensity value (left) and the 2 neighbor feature (right) are shown. Notice how most arrows overlay perfectly on the right image, meaning that the location of the NN found hasn't changed.

2.2. Complexity

The main bottleneck of using the DKL estimator lies in the NN search for all features in the image. A naive brute force approach requires $O(DN_S{}^2)$ operations to find all nearest neighbors in one image. To reduce the computational complexity two main strategies have been followed [10]: Firstly data is stored in a kD tree in which each leaf contains at most L samples as described in [10]. Secondly, the approximate NNs are found using a best-bin first (BBF) bottom-up kD tree traversal. The total complexity of building the tree is $O(DN_S \log(N_S/L))$ and the complexity of an all-NN search is $O(DN_S M)$, where M is a predetermined number of leaf-points to be visited defining the compromise between accuracy and speed.

3. Evaluation

In the present section we evaluate for 2 and 8 neighbors the MI_{DKL} and H_{DKL} alignment criteria's performance against that of standard sampled histogram MI for grayscale images. MRI volumes with known ground truth from a normal brain phantom acquired using Brainweb [1] (available at `http://www.bic.mni.mcgill.ca/brainweb/`) were used.

3.1. Criteria Comparison

When images are registered by iteratively minimizing a dissimilarity criterion it is, in general, difficult to evaluate individually the influence of the optimization and of the criterion. Recently Škerl *et al.* [26] proposed an evaluation protocol to asses the robustness and accuracy of a similarity criterion on a given image pair with known ground truth, *i.e.* "gold standard" registrations, without actually performing any registration. The protocol evaluates the criteria along N randomly oriented line segments passing through the point of correct alignment ($\mathbf{X}_0$) in the K-dimensional parameter space. From these points several measures of the criterion performance can be calculated. Of the five proposed measures we have chosen the three that best describe the accuracy and robustness of the criterion: Accuracy (ACC), Capture Range (CR), or Risk of Non-Convergence (RON):

$$ACC = \sqrt{\frac{1}{N_s}\sum_{n=1}^{N} ||\mathbf{X}_{n,max} - \mathbf{X}_0||^2}$$

$$CR = \min_n(||\mathbf{X}_{n,max} - \mathbf{X}_{n,loc}||)$$

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

ACC is the root-mean square of the distance between the correct alignment X_0 and each of the global maxima $X_{n,max}$ along the $n = 1, 2, ..., N$ sampling lines. CR is defined as the smallest distance between the global maxima and closest minima along each line. RON is the average of positive gradients $d_{n,m}$ within a distance r from each global maxima.

3.2. Tests

We have tested the proposed criteria against MI calculated from standard sampled histograms interpolated using cubic B-Splines for 30 image pairs. The images consist of 10 evenly separated slices along the main symmetry axes (Axial, Sagittal and Coronal) of T1-T2 MRI normal brain volumes simulated by Brainweb [1] with an additive inhomogeneity field $Z(i,j) = \frac{1}{3.2}(i+j)$ to simulate bias (Figure 1). The images are 180x271x180 voxels with slice thickness of 1 mm in all direction, a 9% noise relative to the brightest tissue and an intensity non-uniformity of 40%. The separation between slices was of 5 mm with the first slice corresponding to the center of the axis. The criteria were calculated over a cropped window of fixed size (101x101 pixels) at the center of the images to avoid the influence of the background and changing number of samples.

The tests consisted on rigid transformations composed of a translation and a rotation . The sampling

points were generated from the protocol's authors' website (`http://lit.fe.uni-lj.si/Evaluation/`) using the recommended parameters for the number of lines ($N_s = 50$) and number of sampling points along each line ($M + 1 = 201$). The points are spread over a hyper-shpere with radii 3 degrees and 30 mm (30 pixels). The distance between two consecutive samples is of 0.3 mm and 0.03 degrees.

3.3. Results

The final results are summarized in Table 1. An example of the profiles of the criteria calculated using the images from Figure 1 are presented in Figure 3. It can be seen that the use of the regional features effectively eliminates the strong interpolation artifacts present in the graphs of Figure 1 and that the resulting functions have smooth profiles apt for iterative optimization with H_{DKL} tending to give smoother results than MI_{DKL}. The quantitative results show that all DKL criteria have subpixel accuracy and a large capture range. Entropy calculated using 8 neighbors gives the best overall results. When compared to MI calculated using histograms the criteria calculated from DKL tend to be slightly better in ACC and much better in both RON and CR, indicating a much more robust measure.

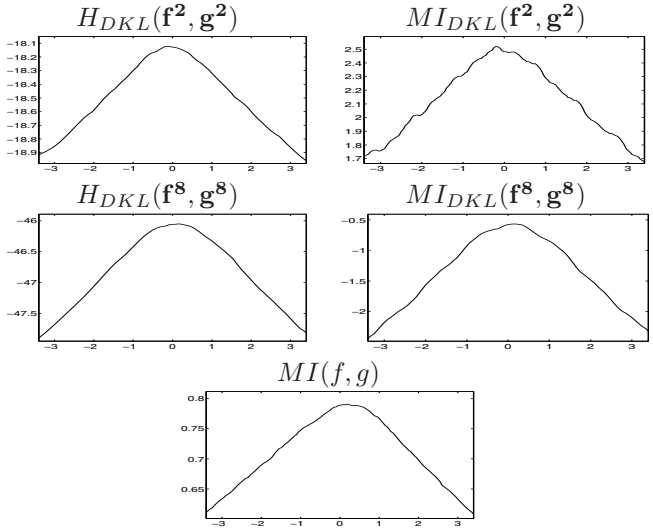

Figure 3. Profiles of the proposed similarity criteria and standard sampled MI vs. displacement in pixels from correct alignment.

4. Applications to color colposcopy images

Ultimately, the quality of an image similarity criterion is determined by its performance within an image registration algorithm. In the present section we will test the robustness of the proposed criterion when used to register cervical colposcopy images. Colposcopy images are acquired during a gynecologic exam shortly after the application of a low concentration acetic acid (3 to 5%) to the cervix. This induces temporal color and texture changes in cancerous and precancerous tissue areas which may be detected by a physician. Registration of these images is the first step of proposed multi-image computer aided diagnostic systems [11].

We have used a sequence of 20 16-bit RGB 562x375 pixel images captured between 60 and 289 seconds after the application of the acetic solution at intervals of 12 seconds. The first and last image of the sequence correspond to the moment of maximum color reaction and to the return to normal coloration, respectively. To measure the registration error we have followed the procedure presented in [6] to semiautomatically generate landmarks from the colposcopy sequence.

4.1. Robustness to displacement

To test the robustness of the proposed criterion against displacement two images from the sequence were aligned rigidly using the extracted landmarks (Figure 4). One of the images was then displaced by a randomly generated starting point composed of a translation of magnitude ρ and a rotation of between ± 5 degrees. Each test consisted of 20 registrations for 16 translation intervals $\rho_l < \rho \leq \rho_{l+1}, l \in [0, 1, 2, ..., 15], \rho_l = 4l$. Registration was done using MI and H_{DKL} with 2 neighbors together with multiresolution processing and a modification of Powell's optimization algorithm [24]. A grayscale version of the images was used to calculate MI from histograms. The registered feature $\mathbf{f^2_{RGB}} \in \mathbb{R}^9$ for the color images is defined as:

$$\mathbf{f^2_{RGB}} = [\mathbf{f^2_R}, \mathbf{f^2_G}, \mathbf{f^2_B}].$$

The 2 neighbor H_{DKL} criterion was chosen over the other DKL criteria to reduce the processing time in each test registration.

All registrations were done over a 200x200 region of interest (ROI) in the template image (seen as a yellow square in Figure 4) roughly centered on the Os region (the end of the uterine cavity) at the center of the cervix. A registration was considered successful if the final transformation had an error of less than 1 pixel in the translation and less than 1 degree in the rotation when compared to the correct transformation. Results are shown in Figure 5.

The color H_{DKL} results in very robust registrations with zero misregistrations for initial displacement values up to 32 pixels and a slow decline in the number of correct registration between 32 and 64 pixels of initial displacement. MI shows good results at very low initial displacements and begins to quickly decline as the distances grow larger. The better results of the H_{DKL} criteria over MI are explained by the lack of strong peaks in the joint histogram and the presence of strong illumination inhomogeneities. For this images most of the alignment information comes

		mean	median	min	max	standard dev.
MI	ACC	0.7191	0.7482	0	1.9442	0.4342
(Cubic B-Spline Interp.)	CR	5.6667	1	1	23	7.2318
	RON	6.4984	6.1403	1.0866	14.555	3.3522
H_{DKL}	ACC	0.4396	0.4472	0	0.9274	0.3121
2 neighbors	CR	14.200	18.000	1	25	8.9420
	RON	1.3034	1.2648	6.5823	2.0100	0.3576
MI_{DKL}	ACC	0.4916	0.4357	0	1.2649	0.3358
2 neighbors	CR	10.800	13.5000	1	23	9.2266
	RON	3.0033	1.6743	6.6497	10.967	2.6885
H_{DKL}	ACC	**0.4008**	0.3464	0	1.1136	0.3415
8 neighbors	CR	**22.6333**	25	1	30.000	7.6180
	RON	**0.5902**	0.5690	0.2293	1.1272	0.1835
MI_{DKL}	ACC	0.4876	0.4345	0	1.1045	0.3367
8 neighbors	CR	14.5677	21	1	28	11.3553
	RON	1.1469	1.1272	6.1709	1.7807	0.3139

Table 1. Results of the criteria comparison protocol. RON is the average of $RON(r) \forall r$,

Figure 4. Images from the colposcopy sequence at $t = 73$ sec. (left) when the acetic reaction is near its maximum and at $t = 289$ sec. (center, registered) when the cervix has already recovered most of its original coloration. A composed image obtained from the registered images is also shown (right). The yellow box corresponds to the ROI for the tests of Section 4.1.

from color and fine features and textures which are lost in the histogram, reducing the effectiveness of histogram estimators and benefiting the performance of the regional features which tend to be unique.

4.2. Sequence Registration

Due to progressive appearance changes of the cervix induced by the acetic acid, changes in the illumination direction and the appearance of new objects in the field of view (*e.g.* glints, mucous, blood spots) the registration between colposcopy images becomes increasingly difficult as the lapse between capture times augments. For small intervals simple similarity criteria such as Sum of Square Differences may be used but will tend to fail when the time interval becomes too large. In the present section we show how the proposed criterion is robust to these appearance changes.

All un-aligned images were registered to the initial image ($t = 60s$) using the method described in 4.1. A ROI separating the cervix from the background was automatically extracted by thresholding the images[14] and applying a sequence of morphological operations to fill holes and isolate the central region. For each test the starting point aligns the centroids of the two ROIs. Results are shown in Figure 6.

As with the tests in section 4.1, H_{DKL} show the best overall results both in accuracy and in robustness. The histogram estimated MI results in several misregistrations, making its use unreliable for robust initial registration in this type of images.

5. Conclusions and Future Work

In the present work we have shown how the registration of difficult medical images may be improved by the

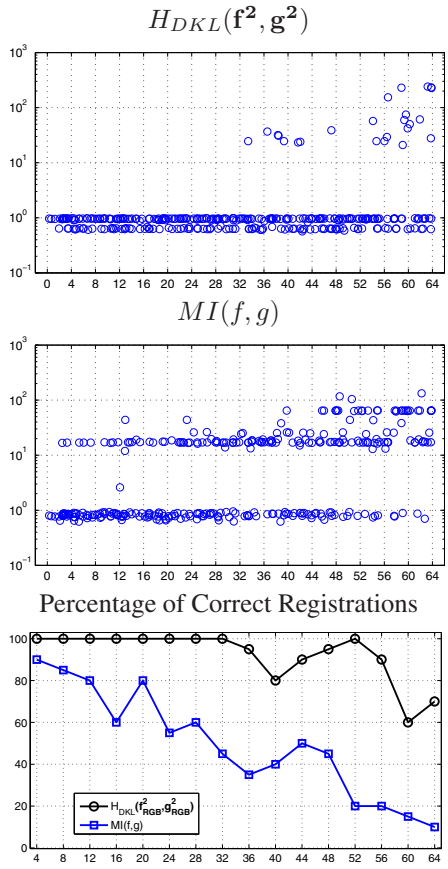

Figure 5. Graphs of initial vs. final displacement position for registrations using $H_{DKL}(\mathbf{f^2})$ (top) and MI (middle). The percentage of correct registrations vs. initial displacement interval is shown for both methods (bottom).

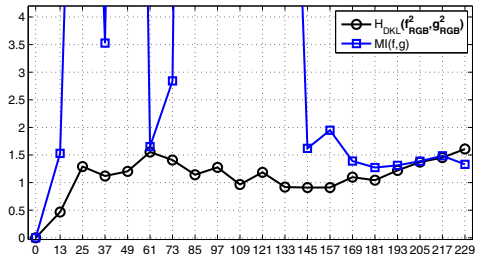

Figure 6. Final error vs. capture time difference between registered images. Large misregistrations using MI data criterion are cropped from the graph.

inclusion of color and high order regional mutual information. Although the use of higher dimensional features may improve registration results in many cases, the lack of adequate pdf estimators has so far prevented their diffusion. We have shown how the KL estimator may be used for this high dimensional estimations in practical way, giving better results than standard MI calculated from histograms.

The main strategy followed to prevent the problems associated with data crashes has been to extract multidimensional features that make each sample unique from each pixel position. This uniqueness makes the distance between samples a better descriptor of the entropy of the image and, therefore, of the shared information between the images. The results suggest that the use of more sophisticated local feature extractors may help in the registration of difficult image pairs.

Finally it should be noted that although only rigid transformations and derivative-free optimization were used in the present article, the use of complex elastic deformations and alternative optimization algorithms with the DKL criteria is possible and remains an open research direction. For this we have derived an analytical expression of the spatial gradient of the H_{DKL} and MI_{DKL} similarity criteria, not presented for lack of space.

6. Acknowledgements

The authors were supported by the Czech Ministry of Education under Projects MSM6840770012 and 1M0567.

References

[1] C. CA, K. Kwan, and R. Evans. Brainweb: online interface to a 3-D MRI simulated brain database. 1997.

[2] J. Chappelow, A. Madabhushi, M. Rosen, J. Tomaszeweski, and M. Feldman. Multimodal image registration of ex vivo 4 Tesla MRI with whole mount histology for prostate cancer detection. In *Medical Imaging 2007: Image Processing. Edited by Pluim, Josien P. W.; Reinhardt, Joseph M.. Proceedings of the SPIE, Volume 6512, pp. 65121S (2007).*, volume 6512 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, Mar. 2007.

[3] H.-m. Chen and P. K. Varshney. Registration of multimodal brain images: some experimental results. In B. V. Dasarathy, editor, *Proc. SPIE Vol. 4731, p. 122-133, Sensor Fusion: Architectures, Algorithms, and Applications VI, Belur V. Dasarathy; Ed.*, volume 4731 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 122–133, Mar. 2002.

[4] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multimodality image registration based on information theory. *Information Processing in Medical Imaging*, pages 263–274, 1995.

[5] R. Gan and A. C. S. Chung. Multi-dimensional mutual information based robust image registration using maximum distance-gradient-magnitude. In G. E.

Christensen and M. Sonka, editors, *IPMI*, volume 3565 of *Lecture Notes in Computer Science*, pages 210–221. Springer, 2005.

[6] J. D. García-Arteaga and J. Kybic. Automatic landmark detection for cervical image registration validation. In M. L. Giger and N. Karssemeijer, editors, *Proceedings of SPIE*, volume 6514 of *Medical Imaging 2007: Computer-Aided Diagnosis*, page 65142S, Bellingham, Washington, USA, February 2007. SPIE, SPIE.

[7] L. F. Kozachenko and N. N. Leonenko. On statistical estimation of entropy of random vector. *Probl. Inf. Trans.*, 23(9), 1987. In Russian.

[8] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Phys Rev E Stat Nonlin Soft Matter Phys*, 69(6 Pt 2), June 2004.

[9] J. Kybic. High-dimensional mutual information estimation for image registration. In *ICIP'04: Proceedings of the 2004 IEEE International Conference on Image Processing*, page 4, 445 Hoes Lane, Piscataway, U.S.A., October 2004. IEEE Computer Society.

[10] J. Kybic. Incremental updating of nearest neighbor-based high-dimensional entropy estimation. In P. Duhamel and L. Vandendorpe, editors, *ICASSP*, pages III–804, 445 Hoes Lane, Piscataway, U.S.A., May 2006. IEEE. DVD proceedings.

[11] H. Lange and D. G. Ferris. Computer-aided-diagnosis (CAD) for colposcopy. *Medical Imaging 2005: Image Processing*, 5747(1):71–84, 2005.

[12] B. Ma, A. Hero, J. Gorman, and O. Michel. Image registration with minimum spanning tree algorithm. In *International Conference on Image Processing*, pages Vol I: 481–484, 2000.

[13] F. Maes, D. Vandermeulen, and P. Suetens. Medical image registration using mutual information. *Proceedings of the IEEE*, 91(10):1699–1722, 2003.

[14] N. Otsu. A threshold selection method from grey-level histograms. *IEEE. Transactions on Systems, Man and Cybernetics*, 9(1):62–66, January 1979.

[15] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Image registration by maximization of combined mutual information and gradient information. In *MICCAI '00: Proceedings of the Third International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 452–461, London, UK, 2000. Springer-Verlag.

[16] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *Medical Imaging, IEEE Transactions on*, 22(8):986–1004, 2003.

[17] D. Rueckert, M. Clarkson, D. Hill, and D. Hawkes. Non-rigid registration using higher order mutual information. In *Proc. SPIE*, volume 3979, pages 438–447, 2000.

[18] D. B. Russakoff, C. Tomasi, T. Rohlfing, and C. R. M. Jr. Image similarity using mutual information of regions. In T. Pajdla and J. Matas, editors, *ECCV (3)*, volume 3023 of *Lecture Notes in Computer Science*, pages 596–607. Springer, 2004.

[19] M. R. Sabuncu. *Entropy-based Methods for Image Registration*. PhD thesis, Princeton University, July 2006.

[20] M. R. Sabuncu and P. J. Ramadge. Spatial information in entropy-based image registration. In J. C. Gee, J. B. A. Maintz, and M. W. Vannier, editors, *WBIR*, volume 2717 of *Lecture Notes in Computer Science*, pages 132–141. Springer, 2003.

[21] C. Shannon. A mathematical theory of communications. *Bell System Tech.*, 27:370–423, 1948.

[22] C. Studholme, D. L. G. Hill, and D. J. Hawkes. Incorporating connected region labelling into automatic image registration using mutual information. In *MMBIA '96: Proceedings of the 1996 Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '96)*, page 23, Washington, DC, USA, 1996. IEEE Computer Society.

[23] D. Tomazevic, B. Likar, and F. Pernus. Multifeature mutual information. In J. M. Fitzpatrick and M. Sonka, editors, *Medical Imaging 2004: Image Processing. Edited by Fitzpatrick, J. Michael; Sonka, Milan. Proceedings of the SPIE, Volume 5370, pp. 143-154 (2004).*, volume 5370 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 143–154, May 2004.

[24] F. Vanden Berghen and H. Bersini. CONDOR, a new parallel, constrained extension of powell's UOBYQA algorithm: Experimental results and comparison with the DFO algorithm. *Journal of Computational and Applied Mathematics*, 181:157–175, September 2005.

[25] P. Viola and W. M. I. Wells. Alignment by maximization of mutual information. *Int. J. Comput. Vision*, 24(2):137–154, 1997.

[26] D. Škerl, B. Likar, and F. Pernuš. A protocol for evaluation of similarity measures for rigid registration. *IEEE T Med Imaging*, 25(6):779–791, 2006.