A novel image similarity measure for image registration

Hrvoje Kalinić and Sven Lončarić
Universiy of Zagreb,
Faculty of Electrical Engineering and Computing,
Department of Electronic Systems and Information Processing,
10000 Zagreb, Croatia
e-mail: hrvoje.kalinic@fer.hr

Bart Bijnens
Catalan Institution for Research and Advanced Studies
and Universitat Pompeu Fabra
E08003 Barcelona, Spain

***Abstract*—The image similarity measure is very important to determine the correspondence between images in order to quantify the accuracy of image registration. The selection of the image similarity measure requires a trade-off between speed and performance. In the current state of the art, fast similarity measures are unable to cope with complex relationships between image intensity values. Currently, the most popular image similarity measures are based on information theory because of their property to find predictable relationship between image intensity values. In this paper, we present a novel image similarity measure and compare it to others such as mean square difference, correlation, correlation coefficient, joint entropy, mutual information, and normalized mutual information. The experiments have shown that the proposed similarity measure is capable of describing complex relationship between image intensity values while offering a favorable speed-performance trade-off as compared to other known similarity measures.**

I. Introduction

An similarity measure is a measure of correspondence between two images. If the similarity measure is maximal, the images are considered to be correctly aligned. Various image registration techniques utilize an image similarity measure to find the correct alignment of two images, for example: multiview, multitemporal or multimodal image registration [1]. The similarity measure may be used for image stitching or mosaicing (multiview), motion detection or object tracking (multitemporal), or for information fusion (multimodal). Considering the problem of aligning two images, the simplest idea is to use the Eucledian distance between images. This idea can be easily extended to different correlation measures which are able to determine the similarity between images with an affine relationship between pixel intensities. However, the assumption of affine relationship between image intensity values does not always hold, especially in the case of multimodal image registration where different sensors are used for image acquisition of the same object, resulting in a complex relationship between pixel intensities. Perhaps for this reason, the most popular similarity measures are those that measure the statistical relationship between pixel intensities, also known as information theoretic similarity measures [2], [3], [4], [5]. A well known example of the information theoretic similarity measures is the mutual information. The mutual information measures how much information one gains about an image given the other image. In this way, the mutual information does not assume any functional relationship between the images, but only a predictable relationship between images [6].

In this paper, we propose a novel image similarity measure and compare its properties to those of other similarity measures. We aim to show that the proposed similarity measure is able to determine the correspondence among images with a functional, rather than just an affine, relationship between their pixel intensities, and this in a relatively short processing time.

II. Similarity measures

The problem of defining the correspondence between images can be framed as a problem of defining a distance measure. The notation used throughout the paper denotes $S(\mathbf{x})$ and $T(\mathbf{x})$ as the source and target image, respectively, where $\mathbf{x}$ stands for the coordinates vector of the image. The deformed source image, or the output image (of the registration process) is denoted as $O(\mathbf{x})$, which can be written as:

$$O(\mathbf{x}) = S(\mathcal{T}(\mathbf{x})) \tag{1}$$

if the geometric transformation of the underlying space of image S is denoted with $\mathcal{T}$. The vector $\mathbf{x}$ is defined on the set D_X defined as either $T \cap O$ [7] or $T \cup O$ [8].

The images O and T can be observed as statistical vectors. Considering the elementary problem of measuring the (dis)similarity between two vectors, the simplest idea is to use a distance measure, for example:

$$D = E[(O(x) - T(x))^2] \tag{2}$$
$$= E[O(x)^2 - 2 \cdot O(x)T(x) + T(x)^2] \tag{3}$$
$$= E[O(x)^2] - 2 \cdot E[O(x)T(x)] + E[T(x)^2] \tag{4}$$

where $E[.]$ denotes the expectation operator over the set D_X. If the $E[O(x)^2]$ and $E[T(x)^2]$ from the equation 4 are constant, the negative of the distance measure D will have qualitatively the same behavior as the correlation between two vectors:

$$C = E[O(x)T(x)] \tag{5}$$

which has a geometric interpretation as the angle between two vectors. For images, the assumption $E[O(x)^2] = E[T(x)^2] =$

const. holds only for small rigid transformations when the overlapping image regions are approximately the same or for the fixed image frame (i.e. $D_X = T \cup O$).

It was shown that the correlation performs well if the images differ only by simple noise (Gaussian [6] or in some case Rician [7]). Various modifications and normalization techniques have been introduced to cope with various noise types or relationships between image pixel values [7], [8], [9]. For example, the correlation coefficient (CC) performs well for an affine relationship between pixel values:

$$CC(T, O) = \frac{E[(T(x) - \mu_T)(O(x) - \mu_O)]}{\sigma_T \cdot \sigma_O} \tag{6}$$

where μ_O and μ_T denotes the image average, and σ_T and σ_O the image standard deviation.

Another approach is to characterize the relationship between T and O in an information theoretic way, by utilizing the joint probability density function (PDF) $p_{TO} = p(T, O)$. The joint PDF allows us to describe any type of functional (or only statistical) relationship between images and any noise that may exists in the image acquisition process. Historically, the first idea to use the joint entropy $H(T, O)$ as an image similarity measure was presented by Hill et al. [10]:

$$H(T, O) = -E[\log p_{TO}(x, y)] \tag{7}$$

Since the joint entropy is defined only on the overlapping region of images T and O (i.e. $T \cap O$), the change in overlap may affect the measured entropy. The solution to this problem was proposed by Collignon et al. [11] and Viola [6], in the form of mutual information (MI):

$$MI(T, O) = E[\log \frac{p_{TO}(x, y)}{p_T(x)p_O(y)}] \tag{8}$$

$$= H(T) + H(O) - H(T, O) \tag{9}$$

Later, Studholme et al. [12] proposed the normalized mutual information (NMI) and showed that this measure, in some cases, performs even better than original MI measure:

$$MI(T, O) = \frac{H(T) + H(O)}{H(T, O)} \tag{10}$$

The latter three similarity measures (H, MI, NMI) are very popular for multimodal image registration due to the property of the joint PDF to describe any type of relationship between images. However, for reliable calculation/estimation of the joint PDF region D_X has to contain a statistically significant number of pixels. Additionally, when compared to the measures defined earlier, they are usually slower to claculate. Therefore, we will briefly discuss the computational complexities of D, C and CC and compare them to H, MI and NMI.

If we use the notation $N = card(D_X)$, where $card(D_X)$ denotes the cardinality of the set D_X, the computational complexity of the CC, C and D is $\mathcal{O}(N)$. Similarly, Maes [13], reported complexity of MI as $\mathcal{O}(N)$, while Roche [14] reported the complexity of $\mathcal{O}(n_x * n_y)$, with n_x and n_y the number of gray levels (or bins) within each image. This difference is relates to whether the histogram or entropy calculation consumes more time. Notice that for the histogram calculation, the algorithm requires all pixel values in the region D_X and the computational complexity is therefore linearly dependent on the number of pixels in the region D_X. However, for the entropy calculation the algorithm has to pass through all bins of the joint histogram (or number of gray value pairs in the images), adding a nonlinear factor to computational complexity. If the number of bins of the histogram is selected to be much smaller than the number of pixels within the image, the MI computational complexity can be approximated by $\mathcal{O}(N)$. Notice that the latter discussion holds not only for MI, but also for H and NMI as well.

Here, we propose a novel exponential correlation (EC) similarity measure of the form:

$$EC = E[(e^{O(x)-\mu_O} - 1)(e^{T(x)-\mu_T} - 1)]^2 \tag{11}$$

where μ_O and μ_T denote the image average. To motivate the form of the proposed EC similarity measure, let us asume the numerator of the Equation 6 (the CC similarity measure) to be the first factor ($i = j = 1$) of the bivariate polynomial of the form:

$$E[\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{ij} \cdot (T(x) - \mu_T)^i (O(x) - \mu_O)^j] \tag{12}$$

With the selection $a_{ij} = \frac{1}{i!} \cdot \frac{1}{j!}$, we can rearrange the Equation 12 to the form:

$$E[\sum_{i=1}^{\infty} \frac{1}{i!}(T(x) - \mu_T)^i \sum_{j=1}^{\infty} \frac{1}{j!}(O(x) - \mu_O)^j] = \tag{13}$$

$$= E[(e^{O(x)-\mu_O} - 1)(e^{T(x)-\mu_T} - 1)] \tag{14}$$

Here, we aim to show that the proposed EC similarity measure from the Equation 11 is able to determine the correspondence among images with complex relationships between the pixel values much better than D, C, or CC. Additionally, we will show that the proposed EC measure can be calculated faster than H, MI, or NMI.

III. Experiments and Results

In this section we compare the accuracy and the execution time of the different similarity measures.

The image database was constructed using the first 200 images returned by a Google search for 512×512 images. After manual discarding the multiple copies of the same image (e.g. several instances of Lena etc.) the testing set was reduced to 167 images. Finally, all images were converted to gray scale. A few example images from the test set are shown in Figure 1.

The accuracy of the image similarity measure is calculated from the correspondence between two aligned images. For this, image registration is utilized. The other set of images for registration is constructed by applying various degradations to the images from the test set. In this way, various changes in the image acquisition process are simulated, assuming the same scene has been imaged. By utilizing the same registration

Fig. 1. A few example images from the test set.

process and changing only the image similarity measure, the registration accuracy will reflect the accuracy of the image similarity measure.

A. Image degradation simulation

During the image acquisition process various factors may affect the quality or change the look of the image. For example, a change in the position of the light source or a change of the sensor type, may significantly change the intensity values within the image. In many image registration applications (e.g. medical image registration) we require that the image similarity measure recognizes the images as similar, regardless of different acquisition conditions, or various changes of image pixel intensity values. Therefore, each image from the set is degraded by random intensity distortion, contrast change, and additive noise. The images degraded in this way are used in the image registration process to test the registration accuracy between the degraded image and the original.

Image degradation effects are implemented in the following way:

1) A uniformly distributed Gaussian noise is added to the image. The amplitude of the added noise is randomly selected from the range [0,A_{max}].
2) The image is degraded by contrast inhomogeneity modeling. The contrast inhomogeneity within an image is modeled via:
$$I'(x, y) = \frac{I(x, y)}{\Delta I(x, y)} \qquad (15)$$
where $\Delta I(x, y)$ is bell-like function defined as:
$$\Delta I(x, y) = k \cdot ((x - x_c)^2 + (y - y_c)^2) \qquad (16)$$
with (x_c, y_c) being the coordinates of the point around which the bell-like curve is positioned and k is constant. The center of the inhomogeniety (x_c, y_c) is randomly selected within the image range, while the factor k is randomly selected from the range [0.00005, 0.0005].
3) To further model a complex relationship between image pixel intensities, a nonlinear intensity distortion is implemented. The distortion is implemented as n-th order polynomial, where both order and roots of the polynomial are randomly selected. The order of the polynomial is selected between values 2 and 6, and roots are selected between A_{min} and A_{max}, where A stands for the intensity level (amplitude). The polynomial is always shifted so that both domain and codomain are [A_{min}, A_{max}] to effectively deform the image intensities.

A few examples of the degraded images can be found in Figure 2.

Fig. 2. Images from the Figure 1 after degradation.

B. Accuracy and precision test

Each image was registered to its degraded version using a different similarity measure. This experiment was independently repeated for image registration utilizing rigid (translation) and non-rigid (scaling) transformation.

For the translation, the similarity measure is calculated for a shift of ± 100 pixels, with a step of one pixel (to avoid interpolation artifact described by Pluim et al. [15]). For the scaling, the scaling factor is calculated for the interval [0.5,2], with a step size of 0.1. In all experiments the similarity measure is calculated in the overlapping image region only, i.e. the set D_X is defined as $T \cap O$, and both translation and scaling are done in the y-axis direction only. The exhaustive search for the maximum, instead of implementing an optimization algorithm, is done to be sure that the global maximum and the correct alignment is achieved.

For the rigid registration (in our case translation only) an accurate registration is achieved for a shift of zero pixels between the original and the degraded image. Therefore, in this case the registration error will be measured in pixels as a misalignment between two images (see Figure 3).

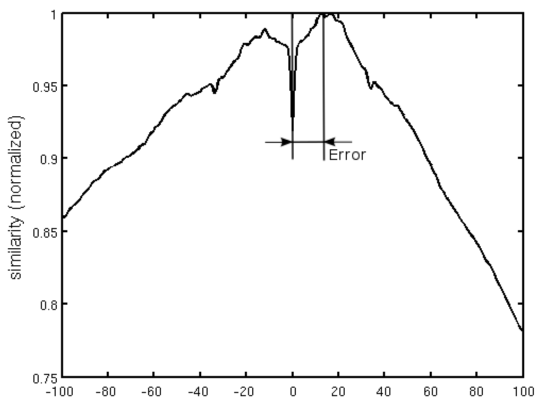

Fig. 3. Similarity measure graph for a translation of ±100 pixels. An incorrect position of the global maximum leads to incorrect registration between original and degraded image. The error is measured in pixels as indicated in on the graph.

The average registration error and standard deviation for rigid registration are show in Figure 4, where circles represents the average error and the lines denotes the distance of one standard deviation from the mean error. The exact numbers are given in Table I. We can notice that the first column (average) reflects the accuracy of the similarity measure, showing a larger bias for some similarity measures (e.g. CC). The second column (standard deviation) reflects the image similarity measure precision, showing that some measures are much more precise than others (e.g. NMI vs. C).

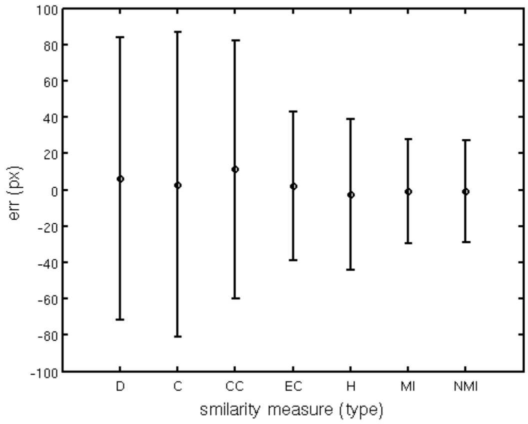

Fig. 4. Average and standard deviation of a registration error (y-axis). The registration utilizes rigid transformation for various types of image similarity measures (x-axis).

	μ_{err}	σ_{err}
D	6.1377	77.7563
C	2.8743	83.9921
CC	11.1557	70.8353
EC	2.0539	40.7180
H	-2.7246	41.5632
MI	-1.0240	28.7559
NMI	-0.8683	28.2774

TABLE I
AVERAGE AND STANDARD DEVIATION OF THE REGISTRATION ERROR. THE REGISTRATION UTILIZES A RIGID TRANSFORMATION FOR VARIOUS TYPES OF IMAGE SIMILARITY MEASURES.

For a non-rigid registration (in our case scaling only) the accurate registration is achieved for a scaling factor of one. In this case the registration error is measured as a unitless value, and calculated as:

$$Err_S = log_2(1 + \epsilon) \quad (17)$$

where ϵ is the absolute error from the accurate registration result (for scaling this should be equal to one). The logarithm is introduced to assure that the scaling error is symmetrical, i.e. it gives the same error for shrinking and stretching the image by the same factor. Also it gives no error if the images are scaled by the same factor. The average registration error and standard deviation for non-rigid registrations are graphically illustrated in Figure 5 using circles and lines, while the exact numbers are listed in Table II.

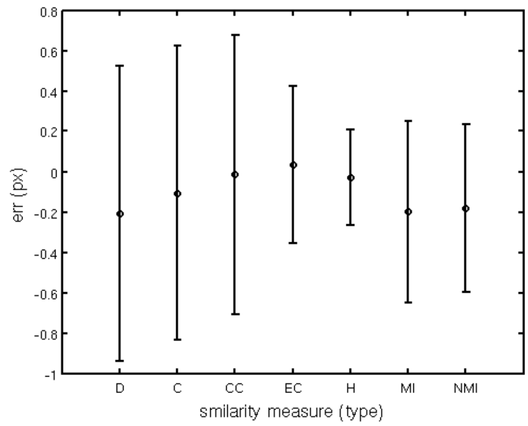

Fig. 5. Average and standard deviation of the registration error (y-axis). The registration utilizes a non-rigid transformation for various types of image similarity measures (x-axis).

	μ_{err}	σ_{err}
D	-0.2077	0.7327
C	-0.1068	0.7285
CC	-0.0147	0.6908
EC	0.0322	0.3890
H	-0.0287	0.2384
MI	-0.1999	0.4517
NMI	-0.1823	0.4139

TABLE II
AVERAGE AND STANDARD DEVIATION OF THE REGISTRATION ERROR. THE REGISTRATION UTILIZES A NON-RIGID TRANSFORMATION FOR VARIOUS TYPES OF IMAGE SIMILARITY MEASURES.

Since the registration procedure is the same for all image similarity measures, the registration error reflects the ability of a similarity measure to recognize images as similar regardless of image degradation effects performed over the image.

From the presented results we notice that in the case of rigid registration, all measures have approximately the same error and all but CC and D have a bias of less than three pixels. However, it is interesting to notice that the standard deviation of the error for EC is lower than D, C, CC and even H, which means that the use of EC as a similarity measure is more reliable. Still, MI and NMI have an even lower error standard deviation. In the case of non-rigid registration, CC, H and EC have the smallest average error, while H and EC have the smallest standard deviation of the error.

C. Speed test

To evaluate the execution time of the similarity measures, the Matlab Profiler was used. The average execution time from 200 function calls is used to compare the performance of the similarity measures. All algorithms were implemented on a standard quad-core PC without parallelization.

The results of the experiment are shown in Figure 6, which presents the average execution time of the different image similarity measures. In Table III the exact numbers for the average execution time, calculated for 200 function calls, are given. From the data we can notice that EC is faster than H, MI and NMI, and almost as fast as CC, but slower than C and D.

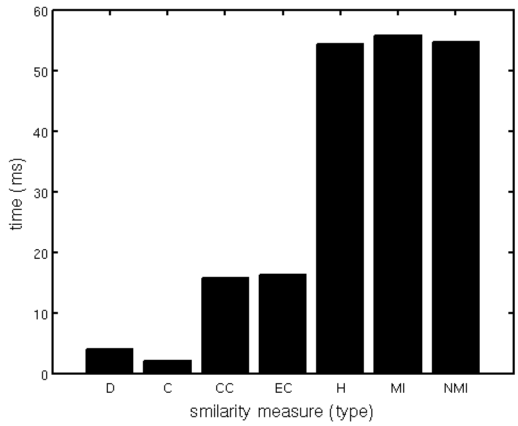

Fig. 6. Comparison of the average execution time of 200 function calls of various image similarity measures.

	$Time$
D	4.020
C	2.030
CC	15.721
EC	16.199
H	54.373
MI	55.647
NMI	54.572

TABLE III
EXECUTION TIME (IN MILISECONDS) FOR EACH SIMILARITY MEASURE.

IV. Discussion and Conclusion

The experiments have shown that the proposed EC image similarity measure performs almost as good as the information theory based similarity measures (H, MI, NMI) for rigid registration, and even outperforms MI and NMI in the case of non-rigid registration. From the presented data one can also notice that the EC outperforms D, C and CC for both rigid and non-rigid registration. Finally, from Section III-C it is evident that although the EC similarity measure is slower than D, C, and CC it is still faster than H, MI, and NMI. The strength of the EC measure is also that it does not require a statistically significant number of pixels for calculation, as opposed to information theoretic image similarity measures. This means that EC can be calculated for even smaller regions where H, MI and NMI do not produce good results. The proposed EC measure inherits calculation speed from the measures such as D, C, and CC while being able to cope with complex intensity relations as done by the information theoretic image similarity measures. From this, we may conclude that the proposed EC similarity measure offers a favorable speed-performance trade-off as compared to other similarity measures described in the paper.

References

[1] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, October 2003.

[2] W. R. Crum, D. L. G. Hill, and D. J. Hawkes, "Information theoretic similarity measures in non-rigid registration," in *IPMI, LNCS 2732*, 2003, pp. 378–387.

[3] J. Zachary and S. S. Iyengar, "Informaton theoretic similarity measures for content based image retrieval," *J. Am. Soc. Inf. Sci. Technol.*, vol. 52, pp. 856–857, August 2001.

[4] L. Cazzanti and M. Gupta, "Information-theoretic and set-theoretic similarity," in *IEEE International Symposium on Information Theory, 2006*, Seattle, WA, USA, 2006, pp. 1836–1840.

[5] D. Lin, "An information-theoretic definition of similarity," in *Proc. 15th International Conf. on Machine Learning*. Morgan Kaufmann, San Francisco, CA, 1998, pp. 296–304.

[6] P. Viola, "Alignment by maximization of mutual information," Ph.D. dissertation, 1995, Co-Supervisors: Lozano-Perez, Tomas and Atkeson, Christopher G.

[7] M. Sonka and J. M. Fitzpatrick, Eds., *Handbook of Medical Imaging*. SPIE Press, 2000, vol. 2. Medical Image Processing and Analysis.

[8] A. Roche, G. Malandain, N. Ayache, and S. Prima, "Towards a Better Comprehension of Similarity Measures Used in Medical Image Registration," in *MICCAI*, 1999, pp. 555–566.

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

[10] D. L. Hill, C. Studholme, and D. J. Hawkes, "Voxel similarity measures for automated image registration," *Visualization in Biomedical Computing 1994*, vol. 2359, no. 1, pp. 205–216, 1994.

[11] A. Collignon and F. Maes and D. Delaere and D. Vandermeulen and P. Suetens and G. Marchal, "Automated multi-modality image registration based on information theory," pp. 263–274, 1995.

[12] C. Studholme, D. Hill, and D. Hawkes, "An overlap invariant entropy measure of 3d medical image alignment," *Pattern Recognition*, vol. 32, no. 1, pp. 71–86, January 1999.

[13] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information." *IEEE Transaction on Medical Imaging*, vol. 16, pp. 187–198, April 1997.

[14] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration," in *MICCAI '98: Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention*. London, UK: Springer-Verlag, 1998, pp. 1115–1124.

[15] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Interpolation artefacts in mutual information-based image registration," *Comput. Vis. Image Underst.*, vol. 77, no. 9, pp. 211–232, 2000.