Kernel-Predictability: A New Information Measure and Its Application to Image Registration

Héctor Fernando Gómez-García, José L. Marroquín, and Johan Van Horebeek

Center for Research in Mathematics (CIMAT),
Apartado Postal 402,
Guanajuato, Gto. 36000, México
{hector, jlm, horebeek}@cimat.mx

Abstract. A new information measure for probability distributions is presented; based on it, a similarity measure between images is derived, which is used for constructing a robust image registration algorithm based on random sampling, similar to classical approaches like mutual information. It is shown that the registration method obtained with the new similarity measure shows a significantly better performance for small sampling sets; this makes it specially suited for the estimation of non-parametric deformation fields, where the estimation of the local transformation is done on small windows. This is confirmed by extensive comparisons using synthetic deformations of real images.

1 Introduction

Image registration is a fundamental task in many fields like medical image processing, analysis of satellital images, and robot vision, among others (see [1][2][3] and references contained there in). Moreover, the methods used to register images, can be adapted to solve other important problems like motion segmentation, stereoscopic registration and the tracking of objects in motion.

When registering two images, I_1 and I_2 , one tries to find the transformation T that applied to I_1 aligns it spatially to I_2. Many registration methods suppose that the intensity of every point is conserved between frames, that is, the equality $I_1[T(x)] = I_2(x)$ is assumed for all the points x; this is known as the *Optical Flow Constraint*, and there is a huge number of registration methods based on this assumption, [4][5][6][7]. However, situations are found very easily where this constraint is violated, for example when the illumination sources change between frames, when the surfaces of the illuminated objects are not lambertian or when registering medical images acquired by different modalities. In these cases, image registration by the maximization of *Mutual Information* (MI) has been widely used because it does not assume a functional relationship between the intensities of the images; instead, it is based on the fact that if aligned, the maximal dependency (information) between the images intensities is found.

A. Leonardis, H. Bischof, and A. Pinz (Eds.): ECCV 2006, Part III, LNCS 3953, pp. 502–513, 2006.

Given two images, I_1 and I_2, their mutual information is defined as:

$$MI(I_1, I_2) = H(I_1) + H(I_2) - H(I_1, I_2) \tag{1}$$

where $H(\cdot)$ is the entropy function defined over the probabilities of the images intensities. For a discrete random variable, the entropy function is written as:

$$H(I) = -\sum_{i=1}^{N} p_i \log p_i \tag{2}$$

with $p_i = p(I = b_i)$, where b_i is the i-th valid intensity value, and for continuous random variables the entropy is written as:

$$H(I) = -\int_{-\infty}^{\infty} p(b) \ln[p(b)] db \; .$$

The first applications of MI to the image registration problem, were published simultaneously by Viola *et al.* [8] and Collignon *et al.* [9], both in the middle of the last decade. Since then, a great number of publications have appeared extending the initial work to problems like nonparametric multimodal image registration [10][11], registration of stereoscopic pairs [12][13] or feature tracking in images [14].

In general, methods based on the maximization of MI, start with an initial transformation T^0, leading to a MI value MI^0, and using a proper optimization method, a sequence of transformations is generated in such a way that the associated MI is increased until convergence. During the optimization process, the increment in MI can be calculated with the expression:

$$\Delta(MI) = \Delta H[I_1(T)] + \Delta H(I_2) - \Delta H[I_1(T), I_2] \; .$$

If the discrete version of the entropy is considered, this is a function of the entries of the probability vector; using a Taylor series expansion, a linear approximation for the increment in entropy is given by:

$$\Delta(H) = -\sum_{i=1}^{N} [1 + \log p_i] \Delta p_i$$

and because the coefficient $[1 + \log p_i]$ is big for small probability values, this increment is highly determined by small features in the images to be registered (which are generally associated with small probability values). This can trap the registration algorithm in local optima, generated when aligning small features, particularly if the small probabilities are not accurately computed. This makes it difficult to apply MI in cases where only a limited sampling is available, for example when measuring entropy at a local level in images, which is important in interesting problems like nonparametric image registration, and in the segmentation of motion between frames, where local measures must be done in order to learn the local motion models and to have enough spatial definition at the motion interfaces.

Another problem related to the application of MI, occurs when working with images with a large background compared to the region of interest, as frequently happens in medical image problems. Under this circumstance the sum of the marginal entropies can become larger than the joint entropy, leading to an increase of MI, instead of decreasing it in misregistration. Studholme *et al.* [15] proposed the use of a normalized version of the MI to overcome this disadvantage. This measure is known as *Normalized Mutual Information* (NMI):

$$NMI(I_1, I_2) = \frac{H(I_1) + H(I_2)}{H(I_1, I_2)} \ . \tag{3}$$

In this work, we propose the use of a new information measure for probability distributions, which we call *Kernel-Predictability* (KP). KP, evaluated in the marginal and joint distributions of two images, is integrated in a similarity measure between images, normalized as (3), and applied to the registration problem. Unlike entropy, the increment of this measure when updated by an optimization method, is mostly determined by the larger entries of the probability vector, which is reflected in a higher robustness in problems where only limited sampling is available. Our proposal is discussed in the next section and in section 3 its performance in image registration problems is compared to that obtained under maximization of MI and NMI. The experimental results show that an important reduction in registration errors is obtained by the use of our method compared to MI and NMI.

2 Kernel-Predictability

In order to introduce our information measure for a given distribution F, consider the following guessing game: someone generates a value x_1 from F and we guess x_1 by generating (independently) a value x_2 from F. We denote by $K(x_1, x_2)$ the obtained reward. More generally, considering various games, we can define the average reward $E(K(X_1, X_2))$. We suppose that the reward function favors guesses close to the true value, i.e., $K(\cdot, \cdot)$ is a decreasing function of the distance between x_1 and x_2. Under this assumption it is clear that the less uncertainty is contained in F, the higher will be the average reward.

The above motivates the following measure for a given distribution F:

$$KP(F) = E[K(X_1, X_2)] = \int_{R^d} \int_{R^d} K(x_1, x_2) dF(x_1) dF(x_2) \ . \tag{4}$$

The last integral is a *regular statistical functional* of degree two (two is the number of arguments in K) [16][17], and the real function K is called the *kernel*. This functional measures the predictability of the random variables distributed according to F, weighted by K, and we denominate it $Kernel - Predictability$.

For the discrete case, this becomes:

$$KP(\mathbf{p}) = \sum_{i=1}^{M} \sum_{j=1}^{M} K_{ij} p_i p_j = \mathbf{p}^T \mathbf{K} \mathbf{p} \tag{5}$$

where M is the number of the different values taken by the random variable X and $\mathbf{K}$ is the matrix which stores in the entry K_{ij}, the reward given for guessing the value x_i when the generated value was x_j. This reward must be maximal, say K_M, when $i = j$, and if it does not depend on i (if we don't have preference to guess any particular value of X), $K_{ii} = K_M$, $\forall i$ and $K_M > K_{ij}$ if $i \neq j$. In general, K_{ij} must be selected as a decreasing function of the distance between the values x_i and x_j.

Observe that if only a unit reward is given for an exact guess, i.e. $\mathbf{K}$ is the identity matrix, KP reduces to the l^2 norm of the probability vector, and $KP(\mathbf{p}) = 1 - G(\mathbf{p})$, where $G(\cdot)$ is the well known *Gini* index of Machine Learning [18]. Opposite to the Gini index and other information measures like the discrete entropy which are invariant under a permutation of the values of the measurement scale, KP can incorporate the quantitative nature of the measurement scale by means of a proper reward function that expresses how close the guess is to the true value.

$KP(\cdot)$ is maximal for random variables which take a fixed value with probability one, by the next inequality:

$$KP(\mathbf{p}) = \sum_i \sum_j K_{ij} p_i p_j \leq K_M \sum_i \sum_j p_i p_j = K_M$$

note that K_M is the value of KP for variables with $p_i = 1$ for any particular value i and $p_j = 0$ for all $j \neq i$.

Taking again $\mathbf{K} = \mathbf{I}$, the minimal KP is obtained for uniformly distributed variables, as can easily be proved. It should be noted that KP is a predictability measure, so it behaves in an inverse way compared to the entropy, which is an uncertainty measure.

Returning to the case $\mathbf{K} = \mathbf{I}$, we can measure the increment in kernel-predictability, which may be associated to the optimization process:

$$\Delta(KP) = 2 \sum_{i=1}^{M} p_i \Delta p_i \ .$$

From this equation one can see that the increment in KP is mainly determined by the larger entries of the probability vector, and for that reason, by the more important features in the images to be registered. This is an important difference with respect to entropy.

2.1 Estimation of the Kernel-Predictability

In practice, it is not always possible to know exactly the distribution function required to evaluate (4) and (5), so an estimation of KP must be done based on a sampling set composed by n independent and identically distributed random variables, $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$,with $X_i \sim F$, $\forall i$. Two estimators to approximate (4) are:

$$\widehat{KP} = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} K(X_i, X_j) \tag{6}$$

$$\widehat{KP'} = \frac{1}{\binom{n}{2}^2} \sum_{i=1}^{n/2} \sum_{j=n/2+1}^{n} K(X_i, X_j) \; . \tag{7}$$

In the first estimator, all possible pairs of variables in $\mathbf{X}$ appear in the sum; in the second, the set $\mathbf{X}$ is divided in two subsets and the kernel is evaluated at each couple formed by taking one variable from the first set and other variable from the second one. Both estimators are unbiased, and if the kernel K is symmetric then $\widehat{KP}$ has the minimal variance among all the unbiased estimators, as shown in [16][17]. $\widehat{KP'}$ has more variance than $\widehat{KP}$ but is cheaper to evaluate. Both variances tend to the same value when the sampling set is increased in size; for these reasons, we use the estimator $\widehat{KP'}$ in the present work.

2.2 Image Registration with Kernel-Predictability

Using KP, one can define a similarity measure between images $I_1(T)$ and I_2, in the following way:

$$SKP(I_1(T), I_2) = \frac{KP[F(I_1(T), I_2)]}{KP[F(I_1(T))] + KP[F(I_2)]} \; . \tag{8}$$

This similarity measure makes a comparison between the predictability of the joint distribution and that of the marginal distributions for the images $I_1(T)$ and I_2. The registration is done by searching for the transformation T^* with maximal SKP, $T^* = \arg max_T[SKP(I_1(T), I_2)]$, due to the fact that the joint distribution of the aligned images gets an ordered and more predictable structure than the one obtained with misregistered images. As done with NMI, our similarity measure is normalized to make it more robust in problems with different content of background and information of interest.

In [19], a measure called Kernel Density Correlation (KDC) is proposed for image registration; that measure shares some similarity with the approximation (7) to the functional (4); however important differences should be noted: firstly, for the approximation of KP (defined in the functional 4), more estimators can be used besides (7), e.g. equation (6), so KP is more general than KDC; moreover in [19] KDC is used for image registration taking the cartesian product of the points in the two images, penalizing the differences in intensities for points that are near in the overlapping coordinates, which implies that this method cannot be applied for multimodal image registration problems; in our case we are using the KP measure to search for peaked joint distributions of intensities in the corresponding regions, which is quite different, and indeed permits its use for multimodal registration.

The similarity defined in (8) can be estimated using the alternatives given above to approximate the KP. In particular, we sample uniformly the coordinates of the images, generating the set $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$; then, evaluate

the intensities of the image I_2 over this set, and the intensities of the image I_1 over the set $T(\mathbf{X})$. Using the estimator (7), the approximation to (8) can be written in the following way:

$$\begin{aligned}\widehat{SKP}(I_1(T), I_2) &= \frac{\sum_{i=1}^{n/2}\sum_{j=n/2+1}^{n} K_2(I_J^i - I_J^j)}{\sum_{i=1}^{n/2}\sum_{j=n/2+1}^{n} K_1(I_T^i - I_T^j) + \sum_{i=1}^{n/2}\sum_{j=n/2+1}^{n} K_1(I_2^i - I_2^j)} \\ &= \frac{\widehat{KP}'_J}{\widehat{KP}'_T + \widehat{KP}'_2}\end{aligned} \tag{9}$$

where $I_J^i = (I_1[T(X_i)], I_2(X_i))$, $I_2^i = I_2(X_i)$, $I_T^i = I_1[T(X_i)]$, K_2 is the kernel used to measure the predictability of the joint distribution of $I_1(T)$ and I_2 and K_1 for the marginal distributions of $I_1(T)$ and I_2. For example, if gaussian kernels are used, then:

$$K_2(I_J^i - I_J^j) = \exp\left\{-\frac{\|I_J^i - I_J^j\|^2}{2\sigma_2^2}\right\} \tag{10}$$

$$K_1(I^i - I^j) = \exp\left\{-\frac{(I^i - I^j)^2}{2\sigma_1^2}\right\} . \tag{11}$$

The maximization can be done using gradient ascent, starting with an initial transformation T^0 and actualizing it with the relation:

$$T^{t+1} = T^t + \lambda \frac{d}{dT^t}\widehat{SKP}(I_1(T^t), I_2)$$

with:

$$\frac{d}{dT}\widehat{SKP}(I_1(T), I_2) = \frac{1}{\widehat{KP}'_T + \widehat{KP}'_2}\frac{d}{dT}\widehat{KP}'_J - \frac{\widehat{KP}'_J}{(\widehat{KP}'_T + \widehat{KP}'_2)^2}\frac{d}{dT}\widehat{KP}'_T$$

and in particular, when using the kernels (10) and (11), the last derivatives are:

$$\begin{aligned}\tfrac{d}{dT}\widehat{KP}'_J &= -\tfrac{1}{\sigma_2^2}\textstyle\sum_{i=1}^{n/2}\sum_{j=n/2+1}^{n}\exp\left\{-\frac{\|I_J^i - I_J^j\|^2}{2\sigma_2^2}\right\}(I_T^i - I_T^j)\tfrac{d}{dT}(I_T^i - I_T^j) \\ \tfrac{d}{dT}\widehat{KP}'_T &= -\tfrac{1}{\sigma_1^2}\textstyle\sum_{i=1}^{n/2}\sum_{j=n/2+1}^{n}\exp\left\{-\frac{\|I_T^i - I_T^j\|^2}{2\sigma_1^2}\right\}(I_T^i - I_T^j)\tfrac{d}{dT}(I_T^i - I_T^j) .\end{aligned}$$

3 Results

This section shows the results obtained by solving the image registration problem, with the application of the similarity measure defined in (9). These results are compared with those obtained by the maximization of MI and NMI. For both versions of mutual information, the entropy was estimated using Parzen windows to approximate the probability densities, following [8]. These approximations are:

$$\begin{aligned}H(I_2) &= -\tfrac{2}{n}\textstyle\sum_{i=1}^{n/2}\ln\left\{\frac{2}{n}\sum_{j=n/2+1}^{n}K_1(I_2^i - I_2^j)\right\} \\ H[I_1(T)] &= -\tfrac{2}{n}\textstyle\sum_{i=1}^{n/2}\ln\left\{\frac{2}{n}\sum_{j=n/2+1}^{n}K_1(I_T^i - I_T^j)\right\} \\ H[I_1(T), I_2] &= -\tfrac{2}{n}\textstyle\sum_{i=1}^{n/2}\ln\left\{\frac{2}{n}\sum_{j=n/2+1}^{n}K_2(I_J^i - I_J^j)\right\} .\end{aligned}$$

The use of Parzen windows is more suitable, when working with limited sampling, than other approaches used to estimate the entropy (e.g. normalized histograms); an additional advantage is that, the approximate similarity measures become differentiable, facilitating the optimization process. Gaussian kernels, (10) and (11), were used to approximate the entropies in MI and NMI (using integration constants to normalize the densities as is required by the Parzen windows) and the corresponding kernel-predictability measure. To make the 3 methods comparable, all the corresponding variances were set to equal values. As is done with the estimation of the KP values, two different, equally sized sampling sets of coordinates are used to estimate the entropy, again following the proposal in [8]. It should be noted that when using Parzen windows the continuous version of the entropy is used; this version can be negative depending on the domain of the variables, and the NMI can be maximal for a negative sum of the marginal entropies and a small negative joint entropy, to avoid this problem the images were scaled to [0,100]. All the methods were optimized using gradient ascent.

3.1 Global Multimodal Image Registration

In the first set of experiments, the 3 methods were tested, using 2 two-dimensional MR images, obtained by the simulator at the Montreal Neurological Institute [20], shown in figure 1. The first image corresponds to a modality T1, with 9% of noise level, and 40% of spatial inhomogeneities in intensity; the second corresponds to a modality T2. A set of 50 random rigid transformations was created, and applied to the T2 MR image. The T1 MR image was used as I_1, so the transformation was always started with the identity. The values for the rotation angles θ were chosen uniformly distributed, $\theta \sim U\{-30°, 30°\}$, and the translation vectors t, taking $t \sim U\{-25, 25\}$ (in pixels). Each sampling set was created taking at random coordinates uniformly distributed in the overlapping region of the images; when this kind of sampling is done, then the part of I_2 which is in the overlap depends on the actual transformation T, and in order to use gradient ascent, the derivative of I_2 with respect to the transformation must be calculated; but I_2 does not depend directly on T, so the gradient of the 3 similarity measures with respect to T was approximated using central finite differences. Though a rigid transformation was applied to the images, the

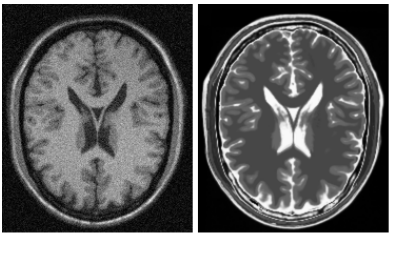


Fig. 1. Images (217×181 pixels) used for the global transformations experiments

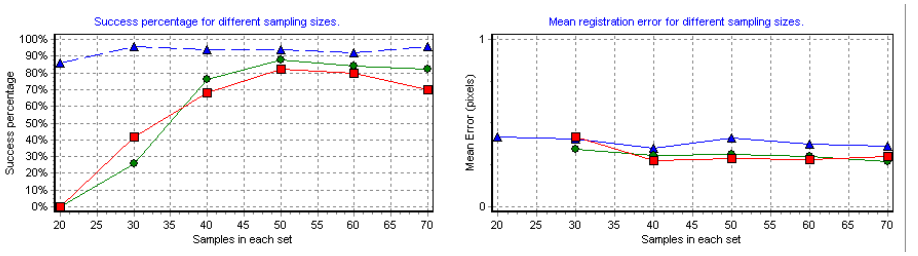


Fig. 2. Success percentage (left) and mean registration error (right) for MI (squares), NMI (circles) and SKP(triangles)

registrations were done searching for the best affine transformation in order to avoid the nonlinearity of the rigid transformations with respect to the rotation angle. In the experiments, the size of the sampling sets was progressively varied, and for each sampling size, the set of 50 registration problems was run.

Figure 2, shows the percentage of successful registrations and the mean registration error obtained for the 3 algorithms. The registration error was measured adding the length of the difference between the applied and the estimated vectors for all the pixels, and then taking the mean value; if the mean value was smaller than 1 pixel, then the registration was considered as successful, and only the successful registrations were considered in the computation of the mean error in the set of 50 transformations. As one can see, one obtains comparable errors using KP, but a significantly higher success rate, specially for small sampling sets, which means that one can obtain performances similar to MI and NMI at a significantly smaller computational cost.

3.2 Local Multimodal Image Registration

The above results suggest that KP should exhibit significantly better performance than MI and NMI when the size of the sampling set is limited by the problem itself; this is the case when the methods are applied for the estimation of nonparametric (local) deformation fields.

To test the performance of the three methods under local multimodal image registration in two-dimensional images, 10 different transformation fields were generated, synthesized by means of two rectangular grids of 5×5 nodes (one grid for each component of the translation vectors) and centering over each node a cubic B-spline function. The nodal values were assigned randomly with values uniformly distributed over a certain interval, in such a way that for each pixel (x, y) a translation vector $[u(x, y), v(x, y)]$ was defined in the following way:

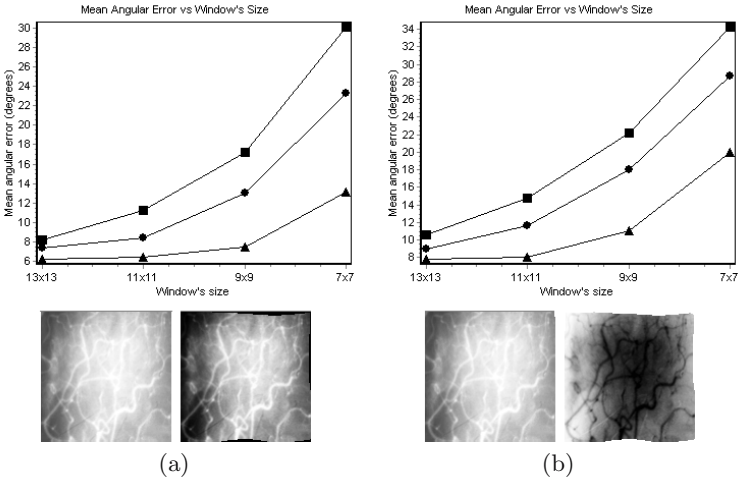
$$\begin{aligned} u(x,y) &= \textstyle\sum_i \sum_j U_{ij} \beta[k_1(x - x_i)] \beta[k_2(y - y_j)] \\ v(x,y) &= \textstyle\sum_i \sum_j V_{ij} \beta[k_1(x - x_i)] \beta[k_2(y - y_j)] \end{aligned} \tag{12}$$

with $U_{ij}, V_{ij} \sim U\{-7, 7\}$, for all centering nodes (x_i, y_j), k_d is the proportion of nodes versus the image dimension in the direction d; and:

$$\beta(z) = \begin{cases} \frac{2}{3} - |z|^2 + \frac{|z|^3}{2}, & |z| < 1. \\ \frac{(2-|z|)^3}{6}, & 1 \leq |z| < 2 \\ 0, & |z| \geq 2. \end{cases}$$

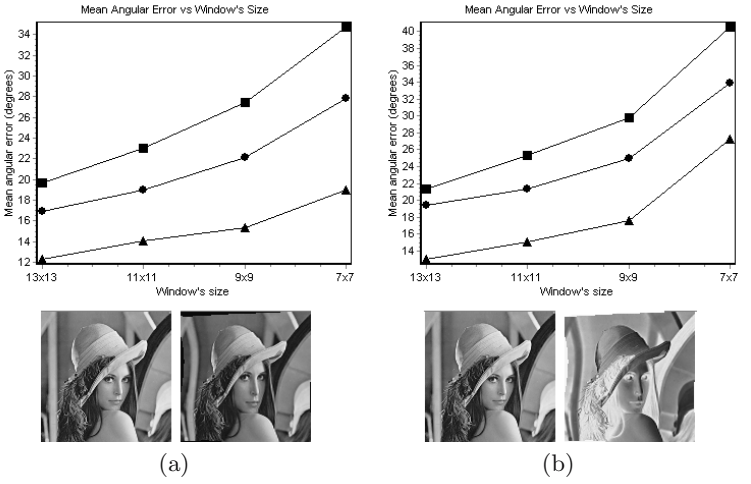
Each generated field, was applied to the images shown in figs. 3 and 4, with 2 different tone transfer functions f_1 and f_2 that distort the intensities of the transformed images. The three registration methods were run locally on a set of nonoverlapping windows centered at pixels uniformly distributed over the images, in order to find the best translation vector that explains the true field $[u(x,y), v(x,y)]$ for each point (x,y); the center points were separated 10 pixels of each other in every direction (the images are 128×128 pixels of size). All the measures were estimated locally, i.e. using only the pixels placed within each window of size $w \times w$. The 2 sampling sets were built by assigning alternatively pixels in the window to each sampling set (each of the sampling sets had $\frac{w^2}{2}$ pixels). The window size was progressively reduced and the registrations repeated for each field. The performance of the three methods was measured using the mean angular error, as proposed by [21]; for that we extend the estimated, $d_e = (u_e, v_e)$, and true vectors $d_t = (u_t, v_t)$ to $d'_e = (u_e, v_e, 1)$ $d'_t = (u_t, v_t, 1)$ now representing the displacement in space and time for every pixel; the angular error is calculated by: $err = \arccos(\frac{d'_e \cdot d'_t}{\|d'_e\| \|d'_t\|})$. Figures 3 and 4, summarize the results obtained by the three methods.

Our registration method can be applied effectively in nonparametric registration problems, as is confirmed in the table 1 which summarizes the results of the application of the 3 measures for the computation of a dense transformation field.



(a) (b)

Fig. 3. Mean angular error for MI (squares), NMI (circles) and SKP(triangles). The subfigure (a) summarizes the results obtained using the tone transfer function $f_1(i) = 100(\frac{i}{100})^{1.35}$, and the subfigure (b) $f_2(i) = 100[1 - (\frac{i}{100})^{1.35}]$, $i \in [0, 100]$. The second row shows the original and the transformed images in each case.



(a) (b)

Fig. 4. Mean angular error for MI (squares), NMI (circles) and SKP(triangles). The subfigure (a) summarizes the results obtained using the tone transfer function $f_1(i) = 100(\frac{i}{100})^{1.35}$, and the subfigure (b) $f_2(i) = 100[1 - (\frac{i}{100})^{1.35}]$, $i \in [0, 100]$. The second row shows the original and the transformed images in each case.

Table 1. Mean angular error for nonparametric registration

Image	SKP	NMI	MI
Figs. 3(a)	13.31 °	17.96 °	20.52 °
Figs. 3(b)	14.89 °	19.06 °	20.97 °
Figs. 4(a)	28.40 °	33.09 °	36.14
Figs. 4(b)	29.19 °	34.29 °	36.27 °

For each of the 3 similarities, the registration was done maximizing the sum of the similarity evaluated in small squared windows centered on each pixel and adding an elastic regularization term. The width of the windows was set to 5 pixels. As was done in the last experiment, 10 synthetic random fields were generated using cubic B-spline functions, but now taking grids with 15×15 nodes for each dimension; each field was applied to the images shown in figs. 3 and 4, with 2 different tone transfer functions, f_1 and f_2. The 3 methods were run for each field and the mean angular error was evaluated.

As can be seen, an important reduction in the registration errors is obtained using our proposal, making it very promising to be applied in local registration problems.

4 Conclusions and Future Work

The similarity measure based in kernel-predictability presented in this paper allows for registrations with errors equivalent to those obtained with MI and NMI but using significantly less sampling. For this reason, our proposal is more

effective for local (nonparametric) registration, based on small windows as is confirmed by the experimental results. The robustness of KP in registration problems with small sampling is due to the fact that the corresponding similarity measure is controlled by the most important features in the images. This robustness makes our measure very promising to be applied in problems where local registration must be done. Our future work will be focused in applying this similarity measure to problems like motion segmentation and image tracking under variable lighting conditions.

Acknowledgement

The authors were partially supported by grant 46270 of CONACyT (Consejo Nacional de Ciencia y Tecnología, México).

References

1. Gottesfeld, L.: A survey of image registration techniques. ACM Computing Surveys, 24(4) (1992) 325–376.
2. Maintz, A., Viergever, M. A.: A survey of medical image registration. Medical Image Analysis, 2(1) (1998) 1-36.
3. Josien, P. W., Pluim, J. B., Maintz, A., Viergever, A.: Mutual information based registration of medical images: a survey. IEEE Transactions on Medical Imaging, 22(8) (2003) 986–1004.
4. Horn B.K.P., Schunck B.G.: Determining optical flow. Artificial Intelligence, 17 (1981) 185–203.
5. Szeliski, R., Coughlan, J.: Spline-based image registration. International Journal of Computer Vision, 22(3) (1997) 199–218.
6. Thirion, J-P.: Image matching as a diffusion process: an analogy with Maxwell's demons. Medical Image Analysis, 2(3) (1998) 243–260.
7. Aubert, G., Deriche, R., Kornprobst, P.: Computing optical flow via variational techniques. SIAM Journal on Applied Mathematics, 60(1) (2000) 156–182.
8. Viola, P., Wells III, W.: Alignment by Maximization of Mutual Information. ICCV, (1995) 16–23.
9. Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G.: Automated multi-modality image registration based on information theory. Information Processing in Medical Imaging. Y. Bizais, C. Barillot, and R. Di Paola, Eds. Dordrecht, The Netherlands. Kluwer, (1995) 263–274.
10. Hermosillo, G., Chefd'hotel, C., Faugeras, O.: Variational methods for multimodal image matching. International Journal of Computer Vision 50(3) (2002) 329–343.
11. DAgostino, E., Maes, F., Vandermeulen, D., Suetens, P.: A viscous fluid model for multimodal image registration using mutual information. MICCAI, (2002) 541–548.
12. Geoffrey, E.: Mutual information as a stereo correspondence measure. Technical Report MS-CIS-00-20, University of Pennsylvania, 2000.
13. Kim, J., Kolmogorov, V., Zabih, R.: Visual correspondence using energy minimization and mutual information. ICCV, (2003) 1033–1040.
14. Dowson, N., Bowden, R.: Metric Mixtures for Mutual Information Tracking. ICPR (2) (2004) 752–756.

15. Studholme, C., Hill, D. L. G., Hawkes, D. J.: An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognit, 32(1) (1999) 71–86.
16. Lee, A. J.: U-Statistics, Theory and Practice. Marcel Dekker Inc. New York, (1990).
17. Lehman, E.L.: Elements of Large Sample Theory. Springer Verlag, New York, (1999).
18. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer Verlag, New York, (2003).
19. Maneesh Singh, Himanshu Arora, Narendra Ahuja.: Robust registration and tracking using kernel density correlation. CVPRW, 11(11) (2004) 174.
20. http://www.bic.mni.mcgill.ca/brainweb/
21. Barron, J. L., Fleet, D. J., Bauchemin, S. S.: Performance of Optical Flow Technics. IJCV, 12(1) (1994) 43–77.