Robust Estimation of Correlation with Applications to Computer Vision

R. Brunelli, S. Messelodi
Istituto per la Ricerca Scientifica e Tecnologica
I-38050 Povo, Trento, ITALY

Abstract— **In this paper we compare to the standard correlation coefficient three estimators of similarity for visual patterns which are based on the L_2 and L_1 norms. The emphasis of the comparison is on the stability of the resulting estimates. Bias, efficiency, normality and robustness are investigated through Monte Carlo simulations in a statistical task, the estimation of the correlation parameter of a binormal distribution. The four estimators are then compared on two pattern recognition tasks: people identification through face recognition and book identification from the cover image. The similarity measures based on the L_1 norm prove to be less sensitive to noise and provide better performance than those based on L_2 norm*.**

Keywords: **template matching, robust statistics, correlation, face recognition, book recognition.**

1. Introduction

The estimation of similarity of patterns is a common low-level vision task which must be routinely performed by many computer vision systems. The Pearson r correlation coefficient is commonly used to compare visual patterns represented as arrays of numbers describing image lightness or related quantities[2]. It is known that the estimation of similarity with the r coefficient is very sensitive to noise, which is not surprising as it is based on the L_2 norm. This paper investigates three alternative correlation estimators, two based on similarity measures derived from the L_1 norm and one on a modified r, which prove to be less sensitive to noise. Fundamental concepts of robustness and related heuristic tools from the literature are used to characterize the statistical properties of the estimators through Monte Carlo simulations. The estimators are compared experimentally in two real world vision tasks: face recognition and book identification.

The first section of the paper presents the similarity measures based on the L_1 norm. A class of robust estimators of position and scale due to Hampel[3] is then used to derive a robust estimator of correlation from the Pearson r. The bias, efficiency, normality and robustness of the estimators are then investigated using heuristic tools and Monte Carlo simulations. Finally, the tasks of face recognition and book identification through template matching are addressed using the described estimators of similarity.

2. Similarity and correlation

The quantification of the similarity of two images is a basic operation for many computer vision algorithms. As visual patterns are usually represented by means of vectors or arrays of numbers, a common choice is to estimate their similarity through the computation of the Euclidean distance:

$$d_2^2(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n}(x_i - y_i)^2 \tag{1}$$

where $\mathbf{x} = \{x_1,\ldots,x_n\}$ and $\mathbf{y} = \{y_1,\ldots,y_n\}$ represent the patterns to be compared. However, images of the same object can be markedly different according to d_2 as a consequence of different ambient illumination or different characteristics of the digitizing device (such as automatic gain and automatic black level adjustment). These differences usually represent a nuisance for many algorithms. To reduce these effects, vectors $\mathbf{x}$ and $\mathbf{y}$ can be normalized in order to have zero average μ and unit variance σ. The distance between the two normalized vectors $\mathbf{x}'$ and $\mathbf{y}'$ can be expressed by:

$$d_2^2(\mathbf{x}',\mathbf{y}') = 2n(1-\rho_{xy}) \tag{2}$$

where ρ_{xy} is the correlation coefficient of the original data. The value of ρ_{xy} represents a similarity measure directly related to the Euclidean distance of the normalized vectors and is commonly used to find out how well a template subimage matches a window of a given image. A major drawback of ρ_{xy} is its sensitivity to even a single grossly erroneous value in the input data:

$$\rho_{xy} \overset{x_i\to\infty,\, y_i = x_i}{\longrightarrow} 1 \tag{3}$$

$$\rho_{xy} \overset{x_i\to\infty,\, y_i = -x_i}{\longrightarrow} -1 \tag{4}$$

This behaviour is related to the properties of the Euclidean distance. The normalization of data is similarly affected by large errors as it is based on the arithmetic mean and on the standard deviation. Alternative *robust* estimators are known in the literature and will be referred to in the following sections. Let us note that the use of a *robust* normalization procedure does not remove the sensitivity of ρ_{xy} to outlying data.

Such erroneous data frequently occur when working with real world images and may be due to transmission noise, specularities, salt and pepper noise, just to cite a few common sources. To cope with the high sensitivity to noise exhibited by ρ_{xy} other similarity measures can be derived from distances other than the Euclidean. An alternative distance, which exhibits a lower sensitivity to noise is the L_1 norm defined by:

$$d_1(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n}|x_i - y_i| \tag{5}$$

Two similarity measures based on the L_1 norm can be introduced:

$$g(\mathbf{x}',\mathbf{y}') = 1 - \frac{\sum_i |x_i' - y_i'|}{\sum_i |x_i'| + |y_i'|} \tag{6}$$

$$l(\mathbf{x}',\mathbf{y}') = \frac{1}{n}\sum_i \left(1 - \frac{|x_i' - y_i'|}{|x_i'| + |y_i'|}\right) \tag{7}$$

*This research was done within MAIA, the integrated Artificial Intelligence project under development at IRST[1]

which satisfy the following relations:

$$\begin{aligned} & g(\mathbf{x}', \mathbf{y}'),\ l(\mathbf{x}', \mathbf{y}') \in [0,1] \\ & g(\mathbf{x}', \mathbf{y}') = 1,\ l(\mathbf{x}', \mathbf{y}') = 1 \quad \Leftrightarrow \quad \mathbf{x}' = \mathbf{y}' \\ & g(\mathbf{x}', \mathbf{y}') = 0,\ l(\mathbf{x}', \mathbf{y}') = 0 \quad \Leftrightarrow \quad \mathbf{x}' = -\mathbf{y}' \end{aligned}$$

where $\mathbf{x}'$ and $\mathbf{y}'$ are normalized vectors. The next section introduces a *robustified* r coefficient. The statistical properties of the estimators are then compared both statistically, in the computation of the correlation parameter of a bivariate normal distribution, and experimentally in two real world vision tasks (face recognition and book identification).

3. Robustness measures of estimators

Whenever a statistical model is applied to real data, low sensitivity of the estimators to slight deviations from the assumed model is highly desirable. Many estimators used routinely, such as the arithmetic mean and the sample standard deviation, are very sensitive to deviations from the models for which they are optimal. The analysis of estimators which are *robust* to deviations from the idealized model is the subject of *robust statistics*[4, 5, 3]. Our analysis will make use of the notion of sensitivity curve[6] of an estimator T based on a finite sample of size n:

$$SC_n(x) = n\left[T\left(\left(1-\frac{1}{n}\right)F_{n-1} + \frac{1}{n}\Delta_x\right) - T(F_{n-1})\right] \qquad (8)$$

where F_{n-1} is the empirical distribution of $(x_1, \ldots, x_{n-1})$. In many situations, $SC_n(x)$ will converge to the influence function[3] of the estimator when $n \to \infty$. The sensitivity curve is mainly a *heuristic tool*, with an important intuitive interpretation: it describes the effect of a small contamination at the point x on the estimate, standardized by the mass of the contamination.

Several robustness measures can be derived from the sensitivity curve of an estimator:

- the *gross-error* sensitivity $\gamma^\star$ which measures the worst approximate influence which a small amount of contamination of fixed size can have on the value of the estimator;
- the *local-shift sensitivity* $\lambda^\star$ which measures the effect of shifting an observation slightly from point x to a neighboring point y;
- the *rejection point* $\rho^\star$ which, when finite, represents the point at which observations are completely rejected.

Another important indicator is the *change-of-variance function* which reflects the sensitivity of the asymptotic variance of the estimator to contamination of the underlying distribution. The related sensitivity parameter is usually named $\kappa^\star$. The above robustness measures are essentially local; a major global robustness measure must be introduced, *the breakdown point*, which describes up to what distance from the model distribution F the estimator still gives some relevant information. A finite sample definition of the breakdown point is the following [3]:

Definition 1: The finite sample breakdown point $\epsilon^\star$ of the estimator T_n at the sample $x_1, \ldots, x_n$ is given by:

$$\begin{aligned} \epsilon^\star(T_n; \mathbf{x}) \quad = \tfrac{1}{n}\max\{k;\ & \max_{i_1,\ldots,i_k} \sup_{y_1,\ldots,y_k} T_n(\mathbf{z}) < \Omega_M, \\ & \min_{i_1,\ldots,i_k} \inf_{y_1,\ldots,y_k} T_n(\mathbf{z}) > \Omega_m\} \end{aligned} \qquad (9)$$

where Ω_M, Ω_m represent boundary values of the parameter to be estimated and the sample $\mathbf{z} = (z_1, \ldots, z_n)$ is obtained by replacing the k data points $(x_{i_1}, \ldots, x_{i_k})$ by the arbitrary values $(y_1, \ldots, y_k)$.

Examples of $\{\Omega_m, \Omega_M\}$ are $\{-\infty, \infty\}$ for location estimators, $\{0, \infty\}$ for scale estimators. An estimator is said to be *robust* if it has a strictly positive breakdown point.

A popular class of robust estimators is given by the M-estimators introduced by Huber [7] generalizing the concept of maximum likelihood estimator. An important example in the class is given by the *tanh*-estimators for scale and location due to Hampel[3]. These operators belong to a particular subset of M-estimators, characterized by the ability to reject extreme outliers completely: their ψ function vanishes outside some central region. The *tanh*-estimators have the following important characteristics:

- a finite rejection point $\rho^\star = r$
- a maximal breakdown point $\epsilon^\star = 0.5$
- a low gross-error sensitivity $\gamma^\star$
- a finite local-shift sensitivity $\lambda^\star$
- a finite change-of-variance sensitivity $\kappa^\star = k$; they also have the highest efficiency subject to the imposed bound $\kappa^\star$ among M-estimators whose ψ-function vanishes outside some central region.

The *tanh*-estimator for scale can be used to construct a robust version of the Pearson correlation estimator r. In fact, for standardized random variables the following formula holds:

$$\rho = \frac{\sigma^2(X+Y) - \sigma^2(X-Y)}{\sigma^2(X+Y) + \sigma^2(X-Y)} \qquad (10)$$

where σ represents the standard deviation. A robust correlation estimator R can thus be obtained by substituting to the standard deviation a robust scale estimator [5] such as the tanh-estimator for a scale parameter:

$$R = \frac{\sigma_{th}^2(X+Y) - \sigma_{th}^2(X-Y)}{\sigma_{th}^2(X+Y) + \sigma_{th}^2(X-Y)} \qquad (11)$$

4. Statistical properties of the estimators

A meaningful comparison of the similarity measures introduced so far requires that they estimate the same parameter. We choose to estimate the correlation parameter ρ of a binormal distribution $b(x, y; 0, 1, \rho)$ with zero averages and unit variances:

$$b(x, y; 0, 1, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)} \qquad (12)$$

A commonly used estimator of ρ is the Pearson coefficient ρ_{xy}. As the average and variance of the marginal distribution of X and of Y are known, they need not be estimated from the data and the Pearson coefficient reduces to r:

$$r(\mathbf{x}, \mathbf{y}) = \frac{\sum_i x_i y_i}{\left(\sum_i x_i^2\right)^{1/2}\left(\sum_i y_i^2\right)^{1/2}} \qquad (13)$$

the denominator being necessary to ensure that $r \in [-1, 1]$. The similarity measures g and l, defined in Eqs. 6 and 7, can be transformed into estimators of ρ. In fact, the similarity measure g is an estimator of

$$E_g = E\left(1 - \frac{|x-y|}{E(|x|+|y|)}\right) = \phi_g(\rho) \qquad (14)$$

while l is an estimator of

$$E_l = E\left(1 - \frac{|x-y|}{|x|+|y|}\right) = \phi_l(\rho) \qquad (15)$$

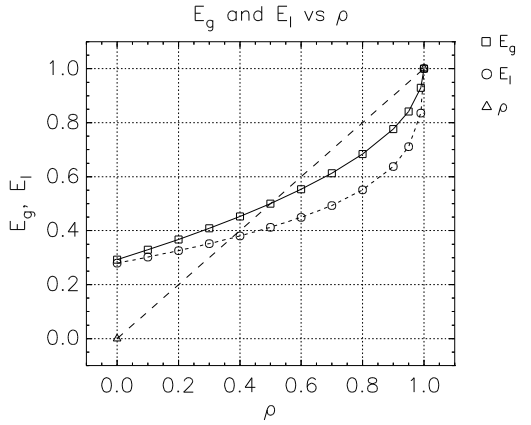

Fig. 1 Plots of $\phi_g(\rho)$ and of $\phi_l(\rho)$ for the bivariate normal distribution $b(0,1,\rho)$

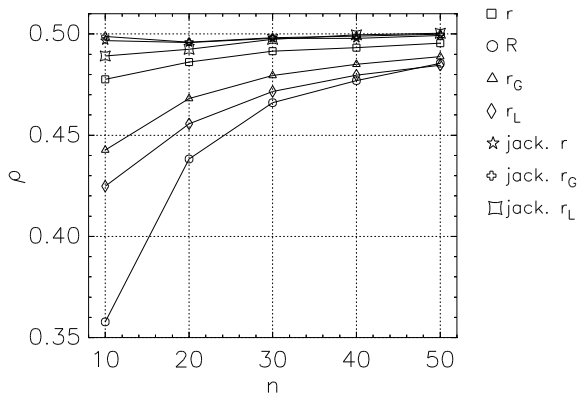

Fig. 2 The biased estimates of the correlation coefficient for $b(0,1,\rho=0.5)$ are reported as a function of the sample size n together with the available jackknifed estimates

where $E(\cdot)$ has the usual meaning of expected value. The mappings ϕ_g and ϕ_l can be inverted and ρ can be computed from E_g and E_l:

$$\rho = \phi_g^{-1}(E_g) \tag{16}$$

$$\rho = \phi_l^{-1}(E_l) \tag{17}$$

When the size of the sample is sufficiently large, g and l are good approximations of the expected values E_g, E_l and can be used as estimators of ρ:

$$r_G = \phi_g^{-1}(g) \tag{18}$$

$$r_L = \phi_l^{-1}(l) \tag{19}$$

The mappings ϕ_g and ϕ_l can be computed numerically under the binormal assumption and are reported in Figure 1.

The estimation of the correlation coefficient of variables distributed according to a bivariate normal distribution falls within the general field of statistical estimation, itself one of the most important methods of statistical inference. Its purpose is to estimate the values of parameters involved in a distribution of a statistical population by using observations on the population. Bias, efficiency, distribution and robustness are important characteristics on which we choose to compare the estimators of correlation R, r, r_G, r_L defined in Eqs. 11, 13, 18 and 19 respectively. The characterization of these estimators makes use of Monte Carlo simulations. In all of the experiments a one-step version of the *tanh* scale estimator was used with $r = 4$ and $k = 6.0$.

4.1 Bias, efficiency and normality

The lower bound given by the Rao-Cramer inequality[8] for the estimation of the correlation of a bivariate normal distribution is:

$$n^{-1}\sigma_Y^{-2} = \int \frac{\rho^4 - 1 + x^2(1+3\rho^2) - 6\rho xy}{(1-\rho^2 1)^3} (y^2(1+3\rho^2) - 2\rho^3 xy)\, db(x,y;0,1,\rho) \tag{20}$$

While the estimators of correlation r, R, r_G, r_L are consistent (i.e asymptotically unbiased) they are all biased and thus the Rao-Cramer inequality does not provide a lower bound for their variance. A popular technique to reduce the bias of an estimator and to estimate its variance is the *jackknife* one, originally proposed by Quenouille [9]. Let us assume that an estimator $\phi(\theta)$ based on a sample of size n has a bias which can be expanded as:

$$E(\phi(\theta) - \theta) = \sum_{k=1} \frac{a_k}{n^k} \tag{21}$$

and that if the i-th observation is neglected the estimate $\phi_i(\theta)$ based on a sample of size $n-1$ has the following bias expansion:

$$E(\phi_i(\theta) - \theta) = \sum_{k=1} \frac{a_k}{(n-1)^k} \tag{22}$$

A new estimator $\tilde{\phi}$ can be introduced:

$$\tilde{\phi}(\theta) = n\phi(\theta) - \frac{n-1}{n}\sum_i \phi_i(\theta) \tag{23}$$

whose bias has no first order term. The so called jackknife *pseudo-values*:

$$\tilde{\phi}_i(\theta) = n\phi(\theta) - (n-1)\phi_i(\theta) \tag{24}$$

can be used to estimate the variance of the estimator ϕ:

$$\sigma^2(\phi) = E[(\tilde{\phi}_i - \tilde{\phi})^2] \tag{25}$$

The jackknife technique is not always applicable and some general conditions are given by Rey[10]. Of the four ρ estimators being compared only R cannot be jackknifed successfully. The biased estimates of the correlation coefficient for $b(0,1,\rho=0.5)$ are reported as a function of the sample size in Figure 2 together with the available jackknifed estimates.

The Rao-Cramer lower bound and the standard deviation of the estimators (jackknifed when possible) are reported in Figure 3 for different values of the parameter ρ.

The knowledge of the distribution of the estimator is of fundamental importance whenever the significance of differences in the estimated values must be quantified. It is known that the distribution of the Fisher's z-transformed Pearson r:

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \tag{26}$$

is approximately normal when the size of the sample is moderately large ($> \quad 10$). The normality of the distribution of the alternative estimators has been investigated using the Kolmogorov-Smirnov test [11]. The significance at which the resulting Fisher z-transformed distributions are normal is reported as a function of the size of the sample in Figure 4.

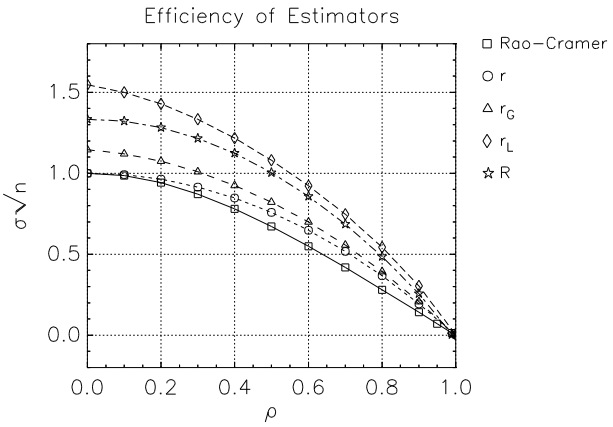

Fig. 3 The efficiency of the estimators as represented by the normalized standard deviation of their distribution at different values of the parameter ρ

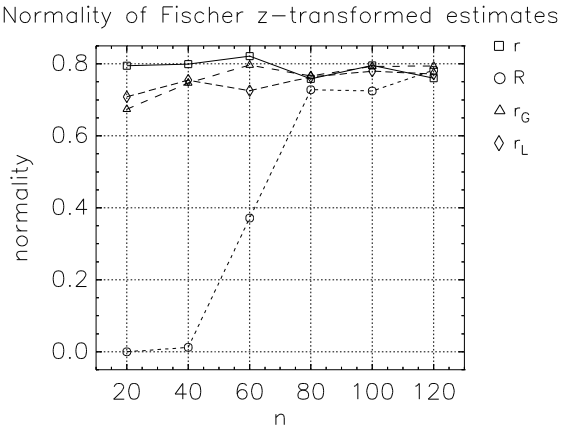

Fig. 4 The normality of the estimators *vs.* the size of the sample as established by the Kolmogorov-Smirnov test on 1000 samples of the estimator distribution at $\rho = 0.5$

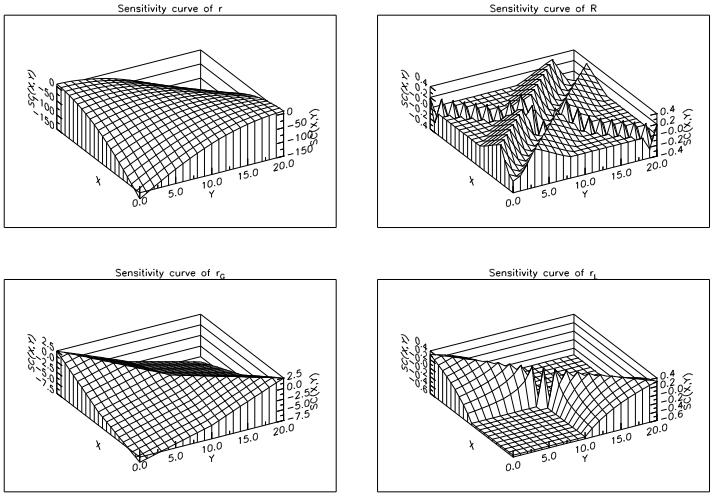

Fig. 5 Sensitivity curves of the r, R, r_G and r_L estimators at $b_{XY}(0, 1, \rho = 0.9)$

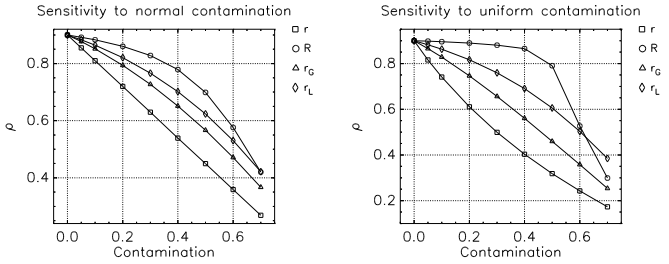

Fig. 6 Sensitivity to the contamination of the bivariate normal ($\rho = 0.9$) with uncorrelated binormal distribution (left) and with an uncorrelated bivariate distribution which is normal on X and uniform on Y (right)

4.2 Robustness

The robustness of the correlation estimators so far considered can be measured by their finite-sample breakdown point defined in Eq. 9. In the estimation of correlation it is appropriate to assume

$$\{\Omega_m, \Omega_M\} = \{-1, 1\} \tag{27}$$

as the set of boundary values in the definition of breakdown point. The breakdown points $\epsilon^\star$ of the different estimators for a sample of size n are:

$$\epsilon^\star_r = 0 \tag{28}$$
$$\epsilon^\star_R = n/(2n) \tag{29}$$
$$\epsilon^\star_{r_G} = 0 \tag{30}$$
$$\epsilon^\star_{r_L} = (n-1)/n \tag{31}$$

so that the only robust estimators are R and r_L. Inspection of the finite sample sensitivity curves SC_n of the estimators, reported in Figure 5, allow us to compare the stability of the estimators: the SC of r_G reflects a superior stability of the estimate to small contamination of the underlying distribution as compared to that of r (though both SC are unbounded). Both R and r_L have bounded sensitivity curves with approximately the same range.

The sensitivity of the estimators to two different contaminations was also investigated through Monte Carlo simulations. The first type of contamination considered was:

$$F' = (1-t)b_{X,Y}(0, 1, \rho) + t\, n_X(0, 1)\, n_Y(0, 1) \tag{32}$$

where the bivariate normal distribution of correlation ρ is contaminated with an uncorrelated bivariate normal distribution; t represents the amount of contamination. The second distribution used in the numerical experiments was:

$$F'' = (1-t)b_{X,Y}(0, 1, \rho) + t\, n_X(0, 1)\, U_Y(-3, 3) \tag{33}$$

where $U_Y(-3, 3)$ represents the uniform distribution in the interval $[-3, 3]$. The sensitivity of the estimators is reported in Figure 6.

The sensitivity of the variance of the estimators under uncorrelated normal contamination is reported in Figure 7. The

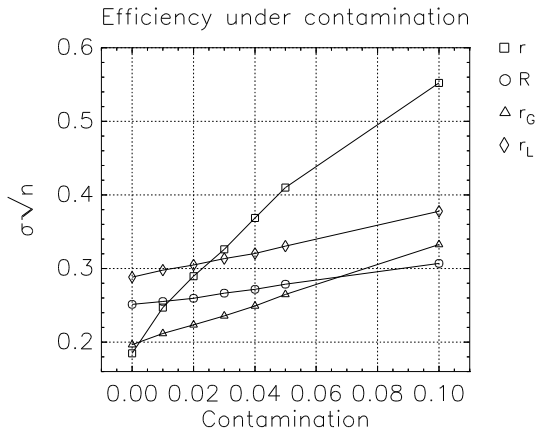

Fig. 7 The efficiency of the estimators as represented by the standard deviation of their distribution under contamination of the bivariate normal with $\rho = 0.9$

Pearson r coefficient exhibits the highest efficiency at the uncontaminated bivariate normal distribution, but even a small contamination makes r the less efficient among the compared estimators.

Note that in all of the Monte Carlo simulations the location and scale of the bivariate normal distribution where not estimated from the sample but were assumed to be known. Should they be estimated, the use of robust estimators (e.g. using the Hampel tanh-estimators) is recommended.

5. Applications

Many visual tasks rely for their successful outcome on the robust estimation of similarity of visual patterns. As an experimental platform for testing computer vision algorithms, we developed an *electronic librarian* whose main goals include updating the electronic catalogue of a library and managing book loans and returns.

In this section we describe the results of the application of the similarity estimators within the two subsystems of the *electronic librarian*: the person identifier and the book recognizer.

5.1 Face recognition

The problem of people identification through face recognition is addressed as a template matching task: the measures of similarity discussed in the previous sections, and the related correlation estimators, provide the matching scores. In the following we will consider images, represented as matrices of numbers, each number representing lightness or a related quantity. The correlation estimators, now to be interpreted as similarity estimators, will be compared in the task of person identification [12, 13]. Digitized images of faces, taken from a frontal view, will be used. The sensitivity of the estimators to three typical kinds of noise which can be expected in digitized images (see Figure 8) is reported in Figures 9,10.

Note that R proves to be the less sensitive estimator while r turns out to be the most sensitive among the compared measures. The estimators were then used to recognize a person by comparing a digitized picture of his/her face against a set of faces stored in a reference database. All of the pictures were geometrically normalized so that the eyes were in the same location [14]. A single set of facial regions, encompassing the eyes, nose and mouth respectively, was used for the comparison (see Figure 11).

Fig. 8 A face image (left) corrupted with additive gaussian, multiplicative gaussian and salt-pepper noise (20% contamination)

Estimator	Recognition (%)	R_{mM}
r	81	1.19
R	81	1.22
r_G	88	1.12
r_L	87	1.18
g	87	1.29
l	87	1.32

TABLE I Performance and separation ratio for the different estimators of similarity. The separation ratio R_{mM} is defined as the similarity of the face to be recognized with the database entry of the corresponding user divided by the highest similarity among the remaining entries.

The corresponding areas of the digitized picture of the unknown face were compared, allowing for limited shifts, to the corresponding areas extracted from the pictures in the database. For each area a correlation measure was computed, thereby obtaining three sets of scores. The similarity of the unknown face to each single face in the database was then computed by using a weighted geometrical average of the three available scores. The resulting performances are reported in Table I.

The r and R coefficient have the same performances, which are lower than those of the r_G and r_L coefficients. The low performance of R, in spite of its robustness, is due to its finite rejection point which, in the task at hand, throws away discriminating information. The high sensitivity of r to even small contaminations of the templates, such as due to a defective geometric normalization of the pictures, is the main reason for its reduced performance. The L_1 related estimators r_G and r_L perform nearly equally well, clearly outperforming r and R. The recognition performance is supplemented by the average value of the ratio R_{mM}, defined as the similarity of the face to be recognized with the database entry of the corresponding user divided by the highest similarity among the remaining entries. The coefficients g and l could also be used to compare the patterns and their performances are reported in Table I. While the recognition performance is essentially unchanged, the value of the R_{mM} ratio is much more favourable and was found to be beneficial when introducing the possibility of rejecting an unknown face because of its lack of similarity with the people in the database. The four estimators have the following computational cost normalized to that of r:

$$\begin{aligned} L_r &= 1.0 \\ L_{r_G} &\sim 1.4 \end{aligned}$$

Estimator	Recognition (%)	R_{mM}
r	86	2.33
g	92	1.50
l	91	1.23

TABLE II Performance of three different similarity indices in a book recognition task. The test set and the database contain respectively 123 and 1003 book

$$L_{r_L} \sim 1.6$$
$$L_R \sim 15$$

The characteristics of robustness, computational efficiency and performance suggest the use of r_L in this particular task.

5.2 Book recognition

The book recognition system aims at recognizing a book by using the image of its cover. The approach adopted in the *electronic librarian* system is based on the selection of a set of global features in order to characterize each book and to distinguish it with respect to a set of candidates[15].

The image of the book is acquired by a module which detects an object entering the field of view of the camera and waits till a stable image is available. The book is then automatically localized and geometrically normalized so that its cover fills a square of predefined size.

The error in the localization of the book vertices ranges from 0 to 20 pixels in a normalized 256x256 image. Besides the localization inaccuracy, the recognition task is complicated by the presence of the user's finger on the book image.

The matching procedure loops on the features list which includes the grey level histogram of the image, the projections of the edges, the location of its most significant peaks and the cover image itself. The comparison stops as soon as a feature allows the classifiers to recognize the unknown book with a predefined confidence level. Further, each feature reduces the number of book candidates inside the data base used for matching the input book.

One of the selected features is the correlation between the cover image of the unknown book and a set of reference images stored in the database.

Figure 12 shows the result of the book acquisition and localization phases, while Figure 13 reports the corresponding normalized database image.

In the presented experiment, 123 book images were acquired and matched against the 1003 database images by using three different similarity indices: r, g and l. The classification performances are compared in Table II.

6. Conclusions

The standard Pearson r coefficient was compared to more stable and/or robust estimators. The high sensitivity of r to even small contamination makes it non suitable to the estimation of similarity of patterns in real world applications. Alternative estimators, based on the L_2 and L_1 norms, were investigated and compared both statistically, in the estimation of the correlation parameter of a bivariate binormal distribution, and experimentally on face recognition and book identification. The similarity measures based on the L_1 norm provide estimators of the correlation parameter of a binormal distribution which are more stable and/or robust of the standard Pearson r coefficient, while retaining a comparable efficiency and computational cost. In the considered real world tasks, the estimators based on the L_1 norm provide reliable scores, outperforming the estimators based on the L_2 norm.

References

1. T. Poggio and L. Stringa. A Project for an Intelligent System: Vision and Learning. *International Journal of Quantum Chemistry*, 42:727–739, 1992.
2. D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice Hall, Englewood Cliffs, NJ, 1982.
3. F. R. Hampel, P. J. Rousseeuw, E. M. Ronchetti, and W. A. Stahel. *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 1986.
4. F. R. Hampel. The influence curve and its role in robust estimation. *J. Amer. Stat. Assoc.*, 69:383–393, 1974.
5. P. J. Huber. *Robust Statistics*. Wiley, 1981.
6. J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977.
7. P. J. Huber. Robust estimation of a location parameter. *Ann. Math. Stat.*, 35:73–101, 1964.
8. R. V. Hogg and A. T. Craig. *Introduction to mathematical statistics*. Collier-Macmillan, 1978.
9. M. H. Quenouille. Note on bias in estimation. *Biometrika*, 43:353–360, 1956.
10. W. J. J. Rey. *Robust Statistical Methods*. Lecture Notes in Mathematics. Springer-Verlag, 1978.
11. J. D. Jobson. *Applied Multivariate Data Analysis*. Springer-Verlag, 1991.
12. R. Brunelli and T. Poggio. Face Recognition: Features versus Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, 1993.
13. L. Stringa. Automatic Face Recognition using Directional Derivatives. Technical Report 9205-04, I.R.S.T, 1991.
14. L. Stringa. Eyes Detection for Face Recognition. *Applied Artificial Intelligence*, 7:365–382, 1993.
15. A. Leonardi, S. Messelodi, and L. Stringa. Book recognition as an example of flat and structured objects classification. Technical Report #9212-07, IRST, Trento, Italy, December 1992. to appear on Cybernetics and Systems.

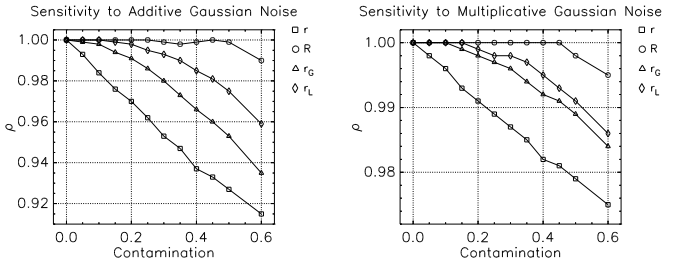

Fig. 9 Sensitivity to additive gaussian noise (left) and to multiplicative gaussian noise such as due to non uniform response of CCD elements in the video camera acquiring the image (right)

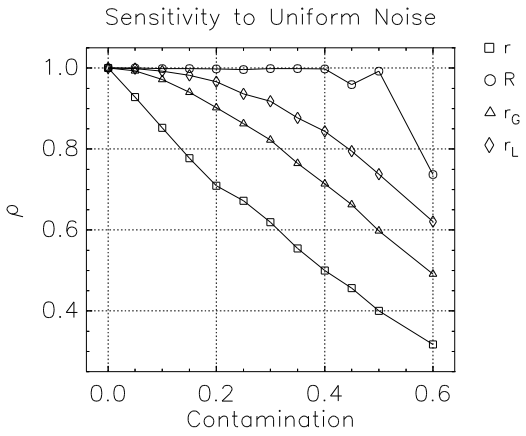

Fig. 10 Sensitivity to salt and pepper noise, which may be due to noise on the transmission channels

Fig. 11 The highlighted regions represent the templates used for identification

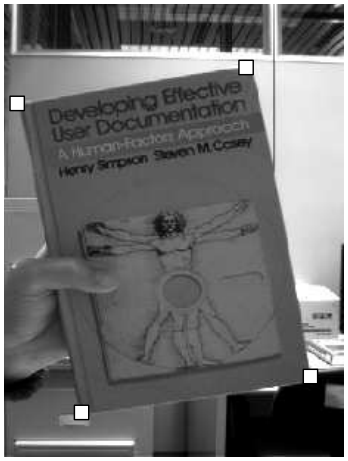

Fig. 12 The image of the book presented to the camera. The vertices of the book are automatically located and are represented in the picture by white squares

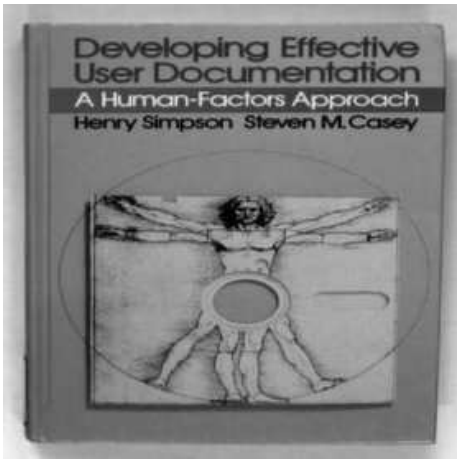

Fig. 13 The reference image stored in the book data base