On the Bin Number Choice of Joint Histogram Estimation Applied to Mutual Information based Face Recognition

Abdenour Hacine-Gharbi[1] and Philippe Ravier[2]
[1]*LMSE Laboratory, Bordj Bou Arreridj University, Elanasser-Bordj Bou Arréridj 34030, Algeria*
[2] *PRISME laboratory, University of Orléans, BP 6744 Orléans Cedex 2, France*

Keywords: Face Recognition, Local Methods, Holistic Methods, Template Matching, Similarity, Mutual Information, Histogram Approach, Bin Number Selection.

Abstract: In this paper, we investigate the binning problem of joint histogram estimation applied to mutual information based face recognition application. Classical approaches for histograms estimation tend to empirically fix the bin numbers. We evaluate in this work some state of the art rules for automatically choosing the bin numbers. The face recognition problem has been studied in the case of local and holistic methods. The choice's performance has been evaluated using AT&T database with single sample in the training set. The results show that better accuracy recognition rates can be achieved with data driven bin number choices rather than fixed bin numbers. In the local method, the results show a higher robustness of the automatic *vs* fixed bin number choice when the regions become smaller.

1 INTRODUCTION

Mutual Information (MI) has been generally used as a measure of statistical dependency between two random variables (Cover and Thomas, 2006) and has therefore become as a popular similarity measure in different signal and image processing applications (Drugman, Gurban and Thiran, October 2007) (Pluim, Maintz and Viergever, 2003). More particularly, it represents the amount of the shared information between random variables.
However the MI computation from data requires the estimation of joint and marginal probability density functions (pdfs) (Cover and Thomas, 2006), which are not known practically and have to be estimated with finite number of samples (Moddemeijer, 1989) (Jain, Duin and Mao, 2000). This estimation can be performed by different approaches such as histogram (Moddemeijer, 1989)(Hacine-Gharbi et al., 2012), Parzen Window (Kwak and Choi, 2002), Gaussian Mixture Models GMM (Ait Kerroum, Hammouch and Aboutajdine, 2010).

The histogram based approach for the mutual information has been extensively used for its undeniable advantages in terms of simplicity and computational complexity (Legg et al., July 2007). In the last decades, this approach has been widely applied in the biomedical (Legg et al., 2013) (Pluim, Maintz and Viergever, 2003) and biometric (Nabatchian, Abdel-Raheem and Ahmadi, 2011) fields as a measure of intensity similarity between images for the image comparison task. In this context, the evaluation of intensity similarity between images is achieved by the joint intensity histogram based mutual information estimation. However, the bin number or equivalently bin width is a crucial parameter that must be carefully selected for the histogram construction in order to avoid a large bias and high mean squared error (MSE) estimation of mutual information (Hacine-Gharbi et al., 2013) (Panzeri et al., 2007). In (Legg et al., July 2007) the authors have used the Sturges' rule in a computer vision problem for selecting the appropriate bin number for histograms estimation. Using the Sturges' rule offers an improvement of the accuracy and efficiency of the registration process when applied on Fundus eye images. In (Legg et al., 2013), the same authors state that there is no definitive answer on the question of how many bins should be used when constructing a histogram.

In the field of facial biometric recognition, a mutual information based method has been proposed in (Makaremi and Ahamdi, 2009). This method locally compares the images by performing the analysis on local regions, separately. For each

Hacine-Gharbi A. and Ravier P..
On the Bin Number Choice of Joint Histogram Estimation Applied to Mutual Information based Face Recognition.
DOI: 10.5220/0004925607770782
In *Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods* (ICPRAM-2014), pages 777-782
ISBN: 978-989-758-018-5

region, the mutual information is computed as a measure of similarity between two images at the same location. In the next, the averaged of the mutual information estimates is considered. The results show that the accuracy rate is dependent of the bin number when chosen as a rule of thumb.

So, in the face recognition literature relating to Mutual Information, most of papers use a fixed number of bins. No rule for selecting the optimal number of bins has been proposed. Hence, in this paper, firstly we investigate the problem of choosing the bin number for face recognition application using different rules. Secondly, we investigate the advantage of a global MI estimation instead of a mean local MI estimation for the application proposed in (Makaremi and Ahamdi, 2009). This is motivated by the fact that previous studies already shown the interest of comparing local and global methods in the field of multimodal image registration (Pluim, Maintz and Viergever, 2003)(Hermosillo and Faugeras, 2001).

2 MUTUAL INFORMATION BETWEEN IMAGES

The Mutual Information (MI) quantifies the information shared between two random variables. Therefore, it is used generally to measure a statistical dependency between variables (Cover and Thomas, 2006). High mutual information means that the variables are very dependent.

For two discrete random variables X and Y, with joint Probability Distribution Function (PDF) $p_{XY}(x,y)$ and marginal distributions $p_X(x)$ and $p_Y(y)$ respectively, the mutual information I(X;Y) of X and Y or simply I_{XY} is defined as:

$$I(X;Y) = \sum_x \sum_y p(x,y) \log\left(\frac{p_{XY}(x,y)}{p_X(x).p_Y(y)}\right) \quad \text{(bit)} \qquad (1)$$

In this paper, we exploit the histogram of intensities of images to estimate the PDFs of facial images. However, bin partitioning in histogram can either be adaptive or uniform (Darbellay and Vajda, 1999) (Hacine-Gharbi et al., 2012). In (Makaremi and Ahamdi, 2009), the authors have chosen a uniform partitioning which gives the better accuracy among a limited set of bin numbers. Hence this requires a high computational cost and cannot guarantee the optimal choice of bin number. To overcome this problem, we propose different rules that have already been proved to be useful in other applications (Legg et al., July 2007)(Hacine-Gharbi et al., 2012).

Let us use an *N*-sample dataset with standard deviation σ. Sturges (ST) proposed a bin number $k = 1 + \log_2(N)$ (Sturges, 1926) while Scott (SC) proposed a bin width $\Delta = 3.5\,\sigma/\sqrt[3]{N}$ (Scott, 1979). Freedman and Diaconis (FD) proposed the bin width computation as $\Delta = 2.IQR(X)/\sqrt[3]{N}$, where the term IQR stands for the interquartile range (Freedman and Diaconis, 1981).

A novel approach for deriving the number of bins has been previously proposed in (Hacine-Gharbi et al., 2012) and (Hacine-Gharbi et al., 2013). In (Hacine-Gharbi et al., 2012), the bin number is chosen in such a way that the bias of MI and entropy estimates is zero. In (Hacine-Gharbi et al., 2013), the optimum number of bins is given by a mean squared error minimization procedure of the estimates. For the two discrete random variables X and Y with standard deviation σ_X and σ_Y respectively, and with extents A_x and A_y respectively, and a correlation coefficient ρ, the Low Mean Square Error (LMSE) strategy for the histogram based estimation of $I(X;Y)$ in eq. (1) gives the following bin number (Hacine-Gharbi et al., 2013):

$$k = \text{round}\left\{\frac{1}{2} + \frac{1}{2}\sqrt{1+4\sqrt{L}}\right\} \quad \text{with} \quad L = \frac{N\rho^2}{12.(1-\rho^2)}\left(\left(\frac{A_x}{\sigma_X}\right)^2 + \left(\frac{A_y}{\sigma_Y}\right)^2\right) \qquad (2)$$

3 ESTIMATION OF MUTUAL INFORMATION BASED ON HOLISTIC AND LOCAL METHODS

Generally, the face recognition methods can be classified into two categories: holistic matching methods and local matching methods. The first holistic approach attempts to identify faces using the whole face region, while other approaches use local regions as inputs to a recognition system (Khan, Javed and Anjum, 2005). In (Ruiz-del-Solar and Navarrete, 2005) comparative studies of different options in a holistic method have been discussed. The options are related to the projection operation of the faces which is helpful for data size reduction, the matching criterion for the similarity measure and the classification method employed. In (Zou and Nagy, 2007), the authors have investigated a comparative study between local matching methods. The methods can be divided into three important steps: 1) alignment and partitioning, 2) feature extraction, and

3) combination / classification. The first step aims at aligning each face into a common coordinate system using affine transforms (translation, rotation and scaling) that produce similar faces. The purpose is to make the classification process insensitive to any variation of the face pause conditions. This step is often difficult. The face partitioning is then achieved for local comparisons. The second step is the feature extraction procedure that consists both of extracting pertinent information from the data and reducing the number of parameters for the classifier. The third step is the classifier that works with the previous input features. The local features can be combined together before classification or considered as input for individual classifiers which results are combined for the final class assignment.

Regarding the last step, measures of similarity between images are mandatory. This task can be achieved using MI measures. Using this tool, the recognition of an unknown facial image can be achieved by searching for the class c (person) that maximizes the MI between the unknown image and each train image. Following this procedure, the authors in (Pluim, Maintz and Viergever, 2003) have mentioned that the MI measure can be estimated globally, on the entire image, or locally, on a sub-image. In particular in (Makaremi and Ahamdi, 2009), the MI has been estimated in local regions. Each image is horizontally divided into two sub-images (left and right sides) with an overlap. Next, each side is divided into smaller overlapping strips. So, each image of class c is represented by a set of strips defined as(Makaremi and Ahamdi, 2009):

$$\mathrm{S}^{\mathrm{C}} = \left\{\mathrm{S}^{\mathrm{c}}_{\mathrm{R,i}}\,, \mathrm{S}^{\mathrm{c}}_{\mathrm{L,i}} \quad \text{for } 1 < i < M\right\},$$

where $\mathrm{S}^{\mathrm{c}}_{\mathrm{R,i}}$ and $\mathrm{S}^{\mathrm{c}}_{\mathrm{L,i}}$ are the ith strips of the right and the left sides of the image respectively and M is the number of strips. Moreover, the mean intensity of each strip is displaced to zero and its histogram is estimated. The MI between an unknown and reference image is estimated as the average of the MI between strips at the similar location. Also mutual information between images with shifts has been considered to avoid comparing strips that not belong to the same face part.

Firstly, we investigate in this paper the choice of bin number to avoid the binning problem in face recognition. Secondly we investigate the MI based recognition system with holistic and local approach. A comparison between the two approaches is therefore carried out with various image partitioning and different parameter value choices in the estimation algorithms.

4 EXPERIMENTAL RESULTS

In this section, we present different experiences in which we discuss the problem of bin number choice for histogram based approach of MI estimation. Moreover, we present from these experiences the performance of holistic and local methods. To achieve these studies, we have chosen the Olivetti Research Laboratory (ORL) database which consists of 40 subjects each having 10 images (The Database of Faces, 2002). This 400 images database has been used in many scientific works as reference database for face recognition (Makaremi and Ahamdi, 2009),(Khan, Javed and Anjum, 2005). The subjects' images have been taken with front view (with tolerance for some side movement), varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses) (The Database of Faces, 2002) (see figure 1).

Figure 1: Examples of images of Faces of 4 subjects (ORL Database (AT&T Laboratories Cambridge,(The Database of Faces, 2002)))

For the construction and testing of the recognition system, one image is used for the training stage and the nine other ones are used for the testing stage. This experience is repeated ten times: each time takes into account the following image in the database for all the subjects as a new training image. The accuracies are then computed as the mean value of ten accuracy estimates.

In the two next sections we discuss the problem of bin number choice in the cases of local method and holistic (global) method respectively.

4.1 Face Recognition Accuracy using Local Method

In order to study the effect of the choice of bin number on system's performance, we have chosen a particular partitioning of the image as shown in figure 2. In a first experience we have divided horizontally the image into two sub-images (left and right sides). Next, each side is divided vertically on M regions (strips) with no overlapping avoiding redundancy between strips.

Once the partition is fixed, the question of the bin number choice that gives the better accuracy now arises. One method consists in testing a bin number choice within a large set of choices and to evaluate the accuracy of the face recognition system for each choice. This technique requires an exhaustive search yielding to high computation costs. To overcome this problem we propose a few bin number estimation rules that exist in literature: Scott (SC), Sturges (ST), Fredman (FD), Low Mean Square Error (LMSE) rules respectively.

Figure 2: Images partitioning.

Table 1 displays the accuracy rates *vs* the bin number (BN) and the number of strips (M) while table 2 displays the accuracy rates considering four rules that estimate the bin number with different numbers of strips M.

Table 1: Accuracy rates as a function of the Number of Bins (NB) and the number of strips (M) in the case of the local method

M \ NB	2	5	10	20	30	50	100
1	56.88	64.00	63.11	63.03	62.39	61.27	55.88
2	53.11	59.06	58.67	58.61	58.06	56.64	49.44
4	51.06	55.97	55.86	54.64	54.75	54.17	42.03
8	47.19	51.31	51.81	51.81	52.28	50.11	25.92
16	43.28	47.00	48.17	48.81	48.72	40.41	18.41

The following remarks and explanations can be made from Table 1:

- Even if optimum values exist in Table 1, the accuracy rate remains approximately the same (with small variations, less than 2 points) when the NB changes from 5 to 30.
- The accuracy of the local method using 2 bins with smaller strips size is better than the method using 100 bins. This can be explained by the insufficient number of samples that cause possible empty bins in the case of 100 bins.
- The accuracy rate decreases when the number of strips M grows, whatever the BN value. This can be firstly explained by the local strategy that greatly suffers from misalignments that may appear between images.
Actually, searching for the optimal alignment is a very difficult task that has not been performed in this study. Secondly, the combination of the local image features (MI between image strips) or the combination of the local classifiers (one classifier per image component or feature) may influence the results. We investigate in this study the average value of the features, whereas other combination exist (Borda counting, majority voting…) (Zou and Nagy, 2007).
- The results show a higher robustness of the automatic *vs* fixed bin number choice when the regions become smaller.

Table 2: Accuracy rates for four bin number estimation rules: SC, ST, FD, LMSE with different numbers of strips (M)

M \ Rule	SC	ST	FD	LMSE
1	64.36	63.06	65.19	64.83
2	61.19	59.06	62.58	61.11
4	57.94	55.81	58.89	59.14
8	54.64	51.69	55.03	54.61
16	51.33	48.17	51.72	51.25

The Table 2 shows the gain that can be obtained with an efficient automatic bin number searching procedure. Indeed, the best accuracy rates given by the automatic estimation of the BN is 3 points superior to the fixed BN case, comparatively, except for the case M=1. The ST method gives the poorer results since this method only takes into account the number of bins N in the BN calculation while the other methods take into account other statistical information about the data (variance for SC, IQR for FD and variance and correlation for LMSE). Actually, the ST method is equivalent to taking a fixed BN for all the histograms to be estimated.

4.2 Face Recognition Accuracy using Holistic Method

Table 3 displays the accuracy rates *vs* the bin number (BN) and the number of strips (M) in the holistic case while table 4 displays the accuracy rates considering four rules that estimate the bin number with different numbers of strips M.

Table 3 shows similar results comparatively to those obtained in the local case (Table 1) with slightly better results when the NB increases for the holistic method.

Table 3: Accuracy rate (ACC) as a function of the Number of Bins (NB) in the case of holistic (global) method.

BN	2	5	10	20	30	50	100
ACC	56.63	63.63	64.36	64.61	64.52	64.19	61.19

Table 4: Accuracy rate for the four bin number estimation rules: SC, ST, FD, LMSE in the case of holistic (global) method

Rule	SC	ST	FD	LMSE
Accuracy rate	66.11	64.52	66.52	65.36

Table 4 shows that the FD method gives the best results with 2 points more with respect to the best result obtained by exploring different fixed BN values displayed in Table 3. This proves that searching for an optimal NB can be attractive practically. The results also show that better accuracy recognition rates can be achieved with data driven bin number choices rather than fixed bin numbers. Indeed, the rule proposed by Sturges only depends on the sample number N and this is equivalent to a fixed bin number since all the images have the same size (112x92 samples). The bin number choice for the other approaches is data driven since the numbers either depend on the standard deviation (SC), the IQR (FD) and the correlation coefficient and standard deviation (LMSE).

Since the LMSE method gives an optimal BN that depends on the correlation parameter between the data (the correlation is not taken into account with the three other methods SC, ST, FD), the accuracy rate for this method should be better since intra class images are highly correlated. However, the LMSE method assumes normal distribution of the data. This hypothesis is not fulfilled with the pixel distribution of the image strips of the full image.

However, it should be noticed that the computational cost for the LMSE method is far lower than the FD method.

5 CONCLUSIONS

In many works in the literature dealing with face recognition, local methods are shown to be preferable than global ones (Hermosillo and Faugeras, 2001). However, many databases need alignment procedures. This task is often difficult to achieve. In this paper, we show that without the alignment procedure, holistic methods are better than local ones with the ORL database using MI based similarity measures.

The automatic BN searching procedure on the data tended to improve the accuracy rate in a recognition faces system based on MI histogram estimations.

The searching procedure is based on criteria which are sensitive to the statistical properties of the data. Adapting the criterion to the exact properties of the data constitutes a perspective of this study.

Other works will be investigated using more “difficult” datasets including effects of illumination variation, pose, and facial expressions. A deeper comparison of local and global approaches will be carried out using such databases (*e.g.* the extended Yale face database B).

REFERENCES

Ait Kerroum, M., Hammouch, A. and Aboutajdine, D. (2010) 'Textural feature selection by joint mutual information based on Gaussian mixture model for multispectral image classification', *Pattern Recognition Letters*, vol. 31, no. 10, July, pp. 1168-1174.

Cover, T. M. and Thomas, J. A. (2006) *Elements of information theory*, 2nd edition, Wiley Series in telecommunications and Signal Processing.

Darbellay, G. A. and Vajda, I. (1999) 'Estimation of the information by an adaptive partitioning of the observations space', *IEEE Transactions on Information Theory*, vol. 45, no. 4, May, pp. 1315–1321.

Drugman, T., Gurban, M. and Thiran, J. P. (October 2007) 'Relevant Feature Selection for Audio-visual Speech recognition', *Proc. of the International Workshop on Multimedia Signal Processing*, Crete, Greece, 179 - 182.

Freedman, D. and Diaconis, P. (1981) 'On the Maximum Deviation Between the Histogram and the Underlying Density', *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 58, no. 2, pp. 139-167.

Hacine-Gharbi, A., Deriche, M., Ravier, P., Harba, R. and Mohamadi, T. (2013) 'New Histogram-based Estimation Technique of Entropy and Mutual Information using Mean Squared Error Minimization', *Computers and Electrical Engineering* , vol. 39, no. 3, pp. 918–933.

Hacine-Gharbi, A., Ravier, P., Harba, R. and Mohamadi, T. (2012) 'Low bias histogram-based estimation of mutual information for feature selection', *Pattern Recognition Letters*, vol. 33, no. 10, pp. 1302–1308.

Hermosillo, G. and Faugeras, O. (2001) 'Dense image matching with global and local statistical criteria: a variational approach', *Computer Vision and Pattern Recognition*, vol. 1, pp. 73–78.

Jain, A. K., Duin, R. P. W. and Mao, J. (2000) 'Statistical Pattern Recognition: A Review', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4-37.

Khan, M. M., Javed, M. Y. and Anjum, M. A. (2005) 'Face Recognition using Sub-Holistic PCA', *First International Conference on Information and Communication Technologies*, ICICT, 152 - 157.

Kwak, N. and Choi, C. (2002) 'Input feature selection by mutual information based on Parzen Window', *IEEE Transactions Pattern Anal Mach Intell*, vol. 24, no. 12, pp. 1667–71.

Legg, P. A., Rosin, P. L., Marshall, D. and Morgan, J. E. (2013) 'Improving accuracy and efficiency of mutual information for multi-modal retinal image registration using adaptive probability density estimation', *Computerized Medical Imaging and Graphics*.

Legg, P. A., Rosin, P. L., Marshall, D. and Morgan, J. E. (July 2007) 'Improving Accuracy and Efficiency of Registration by Mutual Information using Sturges Histogram Rule', *Proc. Medical Image Understanding and Analysis*, Aberystwyth, 26-30.

Makaremi, I. and Ahamdi, M. (2009) 'A Mutual Information Based Face Recognition Method'.

Moddemeijer, R. (1989) 'On estimation entropy and mutual information of continuous distributions', *Signal Processing*, vol. 16, no. 3, pp. 233-248.

Nabatchian, A., Abdel-Raheem, E. and Ahmadi, M. (2011) 'Illumination invariant feature extraction and mutual-information-based local matching for face recognition under illumination variation and occlusion', *Pattern Recognition*, vol. 44, no. 10-11, October, pp. 2576–2587.

Panzeri, S., Senatore, R., Montemurro, M. and Petersen, R. (2007) 'Correcting for the Sampling Bias Problem in Spike Train Information Measures', *Neurophysiology*, vol. 98, no. 3, September, pp. 1064-1072.

Pluim, J. P. W., Maintz, J. B. A. and Viergever, M. A. (2003) 'Mutual Information Based Registration of Medical Images: A Survey', *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, August, pp. 986 - 1004.

Ruiz-del-Solar, J. and Navarrete, P. (2005) 'Eigenspace-Based face recognition: A comparative study of different approaches', *IEEE Transactions Syst.*, vol. 35, no. 3, pp. 315–325.

Scott, D. W. (1979) 'On optimal and data-based histograms', *Biometrika*, vol. 66, no. 3, pp. 605–610.

Sturges, H. (1926) 'The choice of a class-interval', *J. Amer. Statist. Assoc*, vol. 21, pp. 65–66.

The Database of Faces (2002), [Online], Available: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

Zou, J. and Nagy, G. (2007) 'A Comparative Study of Local Matching Approach for Face Recognition', *IEEE Transactions on Image Processing*, vol. 16, no. 10, October, pp. 2617-2628.