Unsupervised band selection for multispectral images using information theory*

J.M. Sotoca, F. Pla, A.C. Klaren
Dept. Llenguatges i Sistemes Informátics. Universitat Jaume I.
Av. Sos Baynat s/n, E-12071 Castellón, Spain
(sotoca,pla,klaren)@lsi.uji.es

Abstract

In this paper, the implication of the relations of information in the case of multispectral images is analyzed. Higher-order mutual information can adopt positive or negative values depending of the correlation among ensembles. Therefore, the existence of negative values reflects higher-order correlations in the conditional informations. On the other hand, the extraction of optimal subsets of spectral images is proposed as a maximization of the conditional entropies at same time that the dependent information among images is minimized.

1. Introduction

In certain application fields where visual information processing is involved, the use of spectral information is of most importance to perform certain tasks, for instance, in remote sensing, fruit quality assessment, etc. The trend for these systems is the use of spectral and spatial information in order to estimate and analyze the presence of chemical compounds, pathologies and other information, providing a qualitative and quantitative evaluation of these features.

Multispectral images are a kind of multimodality, where spectral imaging is combined with digital image processing [5][2]. While the images produced by usual digital cameras contain the intensity, or some colour channels (e.g. RGB), multispectral images provide full spectral information for each pixel over a selected wavelength range. In our experimental work, each spectral image band is recorded for each different wavelength, with a narrow pass-band filter, allowing a multiband representation for each pixel along a given spectral range.

One of the common problems, when having available multispectral data, is how to select the right spectral bands to characterize the problem. Considering multispectral images contain information represented by means of a set of two dimensional signals, the band selection problem in multispectral images could be addressed from the point of view of information theory measures, as some way to deal with relationships among the set of bands representing the multispectral image. Preliminary works in this sense have been applied in pattern recognition using benchmark databases techniques and medical applications[8].

*This work has been supported by grants IST-2001-37306 (IST Project European Union), DPI2001-2956-C02-02 (Spanish CICYT), TIC2003-08496-C04 (Spanish CICYT) and P1-IB2002-07 (Fundació Caixa Castelló-Bancaixa).

2. Connections of Information

Preliminary works in multimodal image registration proposed the use of the co-joint histogram of a set of images, as a frequency space to find a solution for the extraction of sharing information among these images.

In the case of two images, a measure of the information content can be obtained from the co-joint gray level distribution of both images $h(a,b)$. This measure can be obtained by means of the Shannon entropy $H(A,B)$ of the co-joint distribution. The co-joint probability distribution $p(a,b)$ can be estimated from the co-joint histogram $h(a,b)$ as $p(a,b) = h(a,b)/MN$, where the normalizing factor, MN (M columns and N rows), is the number of pixels of one of the images, assuming both images have equal size.

Mutual information $H(A{:}B)$ is a basic concept in information theory [1]. It measures the interdependence between random variables. In the case of two images represented by random variables A and B, the mutual information is defined as [7]:

$$H(A:B) = H(A) + H(B) - H(A,B) \tag{1}$$

where $H(A)$, $H(B)$ are the entropy of images A and B, and $H(A,B)$ is the joint entropy. In the case of discrete random variables, the Shannon entropy $H(A)$ is define as follows:

$$H(A) = -\sum p(a)\log_2 p(a) \tag{2}$$

where $p(a)$ represents the probability distribution of the random variable A.

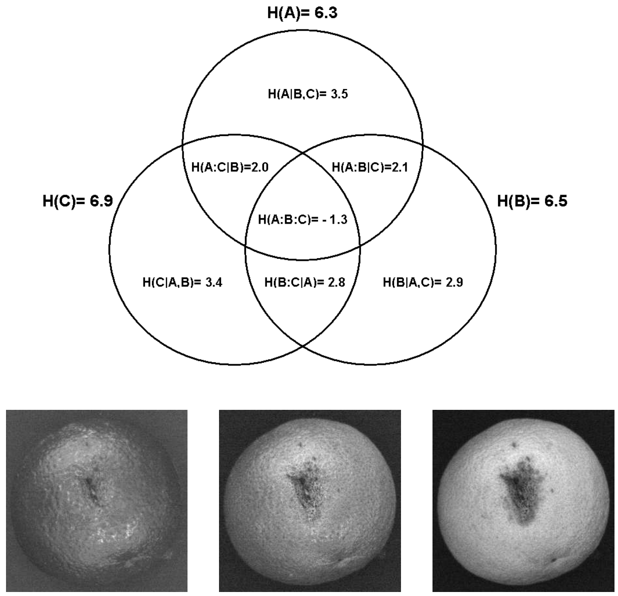


Figure 1. (Top) Entropy Venn diagrams (The circles denote the entropy of an image; joint entropy is the union of circles). (Bottom) Example of spectral image bands: 450, 540 and 580 nanometers.

When $H(A{:}B) = 0$, then the images A and B are independent, i.e., $p(a, b) = p(a)p(b)$. Physically, $H(A{:}B)$ characterizes a measure of the correlation between A and B.

For a three-variable system, we can distinguish the relationship between the random variable A and B due for C into the conditional and mutual information. For instance, the mutual information $H(A{:}B{:}C)$, and the conditional information $H(A{:}B|C)$, can be defined as:

$$\begin{aligned} H(A:B:C) &= H(A:B) - H(A:B|C) \\ H(A:B|C) &= H(A|C) + H(B|C) - H(A,B|C) \end{aligned} \tag{3}$$

where $H(A{:}B|C)$ is the piece of information (between two variables) that is not shared by a third variable, that is, the conditional information on the third variable. Another way to express the mutual information is:

$$\begin{aligned} H(A:B:C) = H(A) + H(B) + H(C) - H(A,B) \\ -H(B,C) - H(A,C) + H(A,B,C) \end{aligned} \tag{4}$$

Relationships among entropies are conveniently represented by entropy Venn diagrams [3][6][9]. In Fig 1, the ternary entropy diagram with variables A, B, C has the following entries: $H(A|B,C)$, $H(B|A,C)$, $H(C|A,B)$ are the conditional entropies; $H(A{:}B|C)$, $H(B{:}C|A)$, $H(A{:}C|B)$ are the conditional informations and $H(A{:}B{:}C)$ is the mutual information. In the case of a classical system, it is known that all the entries are positive, except $H(A{:}B{:}C)$ that can be negative (see Fig 1). So, the system can be characterized by positive or negative values of $H(A{:}B{:}C)$. Therefore, the mutual information $H(A{:}B)$ should contain part of the mutual information $H(A{:}C)$, which implies that $H(A{:}B{:}C) > 0$. Thus, if $H(A{:}B{:}C) < 0$, the relation of pairs of variables are simultaneously unsatisfied. This situation is called *correlation frustration* [6]. The existence of negative values of higher order of mutual information prove the existence of higher-order correlations in the conditional informations. Nevertheless, as we will see later, in our particular problem, we search for extracting a subset of image bands that maximizes the conditional entropies and, at same time, to minimize the dependent information among image bands.

2.1 Information in multispectral images

Let us consider an ensemble of bands $A_1, \ldots, A_n$, where A_i is a random variable of the band i. According to the previous section, let us suppose that the ensemble of bands $A_1, \ldots, A_n$ has a discrete joint probability distribution $p(a_1, \ldots, a_n)$. Then, the joint entropy $H(A_1, \ldots, A_n)$ can be expressed as:

$$H(A_1, \ldots, A_n) = -\sum p(a_1, \ldots, a_n) \log_2 p(a_1, \ldots, a_n) \tag{5}$$

For a single multispetral image, it can be noted that the space of possible events of the joint probability distribution different of zero for a set of spectral bands is fewer or equal that the size of one of the image bands. This implies a reduction of the number of events that take part in the calculation of the entropy. This fact contributes to a reduction of the computing time.

One general expression of the mutual information [6] can be obtained from:

$$H(A_1 : \ldots : A_n) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1 < \ldots < i_k} H(A_{i_1}, \ldots, A_{i_k}) \tag{6}$$

where the sum $\sum H(A_{i_1}, \ldots, A_{i_n})$ runs over all possible combinations $\{i_1, \ldots, i_k\} \in \{1, \ldots, n\}$. Notice that $H(A_1{:}\ldots{:}A_n)$ is symmetric under any permutation of $A_1, \ldots, A_n$. Analogously, a relationship between mutual information in different dimensions can be obtained as Eq 3:

$$\begin{aligned} H(A_1 : \ldots : A_n) = H(A_1 : \ldots : A_{n-1}) \\ -H(A_1 : \ldots : A_{n-1}|A_n) \end{aligned} \tag{7}$$

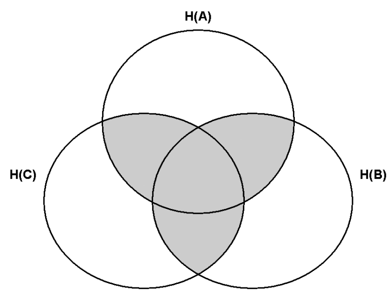
Then, we obtain the following expression for higher-order of mutual information:

$$\begin{aligned} H(A_1 : \ldots : A_n) = H(A_1 : \ldots : A_{n-1}) + H(A_n) \\ + \sum_{k=1}^{n-1} (-1)^k \sum_{(i_1 < \ldots < i_k) \neq n} H(A_{i_1}, \ldots, A_{i_k}, A_n) \end{aligned} \tag{8}$$

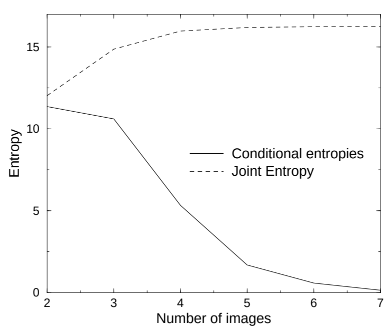
2.2 Estimating discriminant information

An open question is how to define the dependent information among a given subset of image bands. One way could be to measure the mutual information among the image bands. In many problems, as in image registration, the question is how to maximize this dependent information to establish the correspondence between multimodal images [7]. Band selection in multispectral images supposes the opposite problem.

In the band selection problem, we look for a subset of bands that: (a) contains as much information as possible with respect to the whole multispectral image, and (b) the information they represent has to be as much as discriminant as possible, that is, different areas of the images must be represented in such a way the information they represent can be as much as discriminative as possible from other areas representing other image phenomena.



(a)



(b)

Figure 2. (a) Shared information among three image bands. The dark grey region corresponds to the dependent information obtained by Eq 9. (b) Behavior of the sum of Conditional Entropies vs the Joint entropy.

Mutual information measures a certain type of correlation among a set of random variables, but this measure does not represent all possible types of connections of information among them, like all orders of conditional information. On the other hand, conditional entropies provide a measure of independent information content of a random variable given the other ones in the subset. Therefore, a measure of the dependent information of a certain set of random variables could be defined through the following expression:

$$\Theta_{DI} = H(A_1, \ldots, A_n) - \sum_{i_1=1}^{n} H(A_{i_1} | A_{i_2}, \ldots, A_{i_n}) \quad (9)$$

where $A_{i_2}, \ldots, A_{i_n}$ are the complementary variables of A_{i_1}. In the case of a three bands image (see Fig 2(a)), the dark grey region would correspond to such a measure.

Given certain number of bands, the selection of a subset of bands from a multispectral image that minimizes the above criterion would provide a set of bands with minimum interdependence, trying to keep as much information content as possible. This criterion will be called hereafter the *Minimization of the Dependent Information* (MDI).

So, when a certain number of bands is searched, a simple strategy to look for a subset that minimizes the MDI criterion consists of a forward approach, beginning from the image with the highest entropy and adding in a sequential forward scheme the image bands that obtain a biggest reduction of the MDI measure described in Eq 9.

When increasing the number of image bands during the process, the joint entropy keeps growing, increasing the information content. On the other hand, the sum of their conditional entropies decrease with respect to the joint entropy, while trying to keep the selected bands as independent as possible.

Notice that MDI criterion is definite positive, since it always holds that $H(A_1, \ldots, A_n) >= \sum_{i_1=1}^{n} H(A_{i_1} | A_{i_2}, \ldots, A_{i_n})$, therefore, when MDI criterion is null, the random variables are fully independent among them.

3 Empirical Results

The collection of multispectral images used in the experiments were obtained by an imaging spectrograph (Retiga-Ex, Opto-knowledged Systems Inc., Canada). The spectral range extended from 400 to 720 nanometers in the visible (VIS), with a spectral resolution of 10 nanometers, obtaining a set of 33 spectral bands for each image.

The image database consisted of oranges with different types of defects on their surface. From the 33-band images, the MDI criterion was used to select the subset of a given number of bands with higher relevance to discriminate and detect the different types of defects in the fruits.

In order to assess the performance of the method, a Nearest Neighbor (NN) classifier was used to classify pixels into different classes of defects and fruit surface. A set of 21684 pixels as representative examples of the different regions of the oranges were labeled into five classes to build a training set. Each pixel represented is a pattern defined by a feature vector consisting of the values of the pixel in each spectral band. One of the classes represented to the healthy orange skin and the other four classes belong to different typologies of defects.

To increase statistical significance of the results with a limited number of instances, 5-fold cross-validation was employed in the data set. So, the results reported here correspond to the average over the five partitions.

Apart from the unsupervised approach here proposed using the MDI criterion, a supervised filter feature selection method was used. Thus, the band selection method was considered as a supervised feature selection approach. In this case, the well-known *ReliefF* algorithm [4] was used because of its widespread use and good performance in general feature selection problems. This method assigns higher weights when more relevant is a feature.

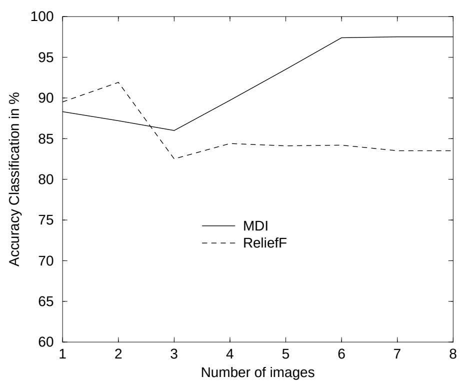


Figure 3. Accuracy of the *NN* rule using the selected bands obtained with MDI and *ReliefF*.

In the Fig 3, we can notice that the two first bands chosen by *ReliefF* improve the accuracy of the classifier with respect to MDI. However, it is not clear to what extend the *ReliefF* scheme can detect redundant or highly interacting attributes. Thus, for this type of problems, where there exists high correlation among different bands with common information, which could be represented with a smaller number of bands, selection of non correlated and high content information is necessary to obtain a good accuracy in the classification. Therefore, the MDI criterion with only six images reaches a classification performance of 97.4%, while using the 33 images of the VIS range, the classification performance is 98.5%. Notice also the difference in the performance between the MDI approach and the supervised feature selection method used.

Note also that from the subset of six images, the improvement of the classification accuracy is not significative. This is due to the fact that independent information present in the subset chosen is near the maximum joint information content that can be reached, so any new image band contributes very little to the conditional entropy term in the MDI criterion.

4 Conclusions and Future Work

In this paper, a study has been realized about higher-order connections of information among bands in multispectral images. From the information theory point of view, different properties can be measured for multispectral images to know about their relationships and information content. The extraction of optimal subsets of spectral image bands can be obtained by means of criterion based on the proposed Minimization of the Dependent Information (MDI), which consists of a relation between the joint entropy and the union of the conditional entropies. This criterion looks for sets of spectral bands with minimum interdependency and high information content.

The principal advantage of this technique is its unsupervised nature, against other supervised techniques like feature selection approaches, providing also better performance, from the classification point of view, than classical filter feature selection approaches like *ReliefF* .

References

[1] J. Aczél and Z. Daróczy. *On measures of information and their characterization*. Academic Press, New York, 1975.

[2] G. H. Bearman, R. M. Levenson, and D. Cabib. *Spectral imaging: Instrumentation, applications and analysis*. SPIE Publications, 2000.

[3] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons Inc., New York, 1991.

[4] I. Kononenko. Estimating attributes: analysis and extensions of relief. In *Proc. of 7th European Conf. on Machine Learning*, pages 171–182, Catania, Italy: Springer Verlag, 1994.

[5] R. Leitner, H. Mairer, and A. Kercek. Real-time classification of polymers with nir spectral imaging and blob analysis. *Real-Time Imaging*, 9(4):245–251, 2003.

[6] H. Matsuda. Physical nature of higher-order mutual information: Intrinsic correlations and frustration. *Physics Review E*, 62(3):3096–3102, 2000.

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

[8] G. D. Tourssari, E. D. Frederick, M. K. Markey, and C. E. J. Floyd. Applications of the mutual information criterion for feature selection in computer-aided diagnosis. *Medical Physics*, 28(12):2394–2402, 2001.

[9] T. Tsujishita. On triple mutual information. *Advances in Applied Mathematics*, 16(1):269–274, 1995.