Biomedical Signal Processing and Control 20 (2015) 16–23

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Segmentation of multispectral images based on band selection by including texture and mutual information

Riad Khelifi*, Mouloud Adel, Salah Bourennane

Aix-Marseille Université, CNRS, Centrale Marseille, Institut Fresnel UMR 7249, 13013 Marseille, France

ARTICLE INFO

Article history:
Received 24 April 2014
Received in revised form 28 January 2015
Accepted 28 January 2015
Available online 25 April 2015

Keywords:
Minimization of the dependent information
Multispectral band selection
Information theory
Joint entropy
Conditional entropy
Dependent texture information
Co-occurrence matrix

ABSTRACT

Minimization of the dependent information (MDI) has been widely used for the selection of multispectral bands. It selects multispectral image bands using information theory concepts that consist of a relation between the joint entropy and the union of the conditional entropies of the considered set of image bands. In this paper, a new band selection method is proposed for segmentation of multispectral textures such that relevant texture information is maximized while reducing the number of spectral bands. Therefore, the proposed method could be viewed as an enhancement of MDI method. The idea is based on MDI, by using the co-occurrence matrices, under Classification Accuracy Rate (CAR) and computational requirement constraints. According to the optimal band subset selected by the proposed method, we make a segmentation of multispectral images and evaluate its accuracy. Experimental results show that the proposed band selection method compares favorably with the MDI.

1. Introduction

Multispectral sensors collect information as a set of images. Each image represents a wavelength of the electromagnetic spectrum and is also known as a spectral band. These images are then combined and form a three-dimensional multispectral data cube for processing and analysis. In order to exploit multispectral imagery in applications requiring high spatial resolution, e.g., urban land-cover mapping, crops and vegetation mapping, tissues structure identification, it is necessary to incorporate spatial [1], contextual [2] and texture information in the multispectral image classification and segmentation processes. Many studies have been dedicated to multispectral texture segmentation or classification [3,4] and a lot of methods have been published for 2-D grey level images [5–9]. However, the main problem of texture analysis of multiband images is related to the high dimension of the data and its high correlation. Driven by classification or discrimination accuracy, one would expect that as the number of multi or hyperspectral bands increases, the accuracy of classification should also increase. Nonetheless, this is not the case in a model-based analysis [10,11]. Redundancy in data can cause convergence instability of models. Furthermore, variations due to noise in redundant data propagate through a classification or discrimination model. Thus, processing a large number of multi or hyperspectral bands can result in higher classification inaccuracy than processing a subset of relevant bands without redundancy [11]. One way of overcoming this problem is to adopt a proper selection band method before applying classification task. The reason is that, in a selection band procedure, the amount of data is reduced into a lower dimensional subspace without practically losing relevant information [12].In addition, computational requirements for processing large hyperspectral data sets might be prohibitive and a method for selecting a data subset is therefore sought [10]. Several approaches have been investigated by looking into how to remove information redundancy either by feature extraction or feature selection resulting from highly correlated bands [10,13–20]. Most of the methods usually involve two separate tasks: (a) selecting the bands that can indicate the particular material well, feature bands selection; and (b) removing the feature bands contributing redundant information, redundancy reduction [21]. Information theory has also been used in feature bands selection [22,20]. It consists in analyzing the amount of information in a subset of features (bands), measuring the degree of independence between image bands as a relevance criterion.

In this paper the main objective is to enhance the MDI method proposed by [20]. Therefore, we are interested in selecting an optimal set of spectral bands for segmenting multispectral images of prostate tissue.

* Corresponding author. Tel.: +33 619327577.
E-mail addresses: riad.khelifi@fresnel.fr (R. Khelifi), mouloud.adel@fresnel.fr (M. Adel), salah.bourennane@fresnel.fr (S. Bourennane).

http://dx.doi.org/10.1016/j.bspc.2015.01.009
1746-8094/

For this purpose, we explore the idea of band selection using both texture and mutual information for the problem of segmenting multispectal images already addressed in [23–25]. This can be achieved by minimizing the dependent information between co-occurrence matrices of multispectral image bands.

The remainder of this paper is organized as follows: Section 2 presents an overview of co-occurrence matrices procedure. Section 3 makes a link between mutual and texture information for selecting the best relevant spectral bands. Section 4 gives more details about the proposed band selection strategy. Section 5 provides some comparative results between our proposed method based on the minimization of the dependent texture information (DTI) and the MDI approach [20]. The goal of this section is to consider the segmentation accuracy as an evaluation criterion in order to compare the influence of the selected image bands obtained by DTI and MDI methods. Section 6 concludes this paper.

2. Co-occurrence matrices for texture segmentation

The most known procedure for extracting textural properties of image in the spatial domain is the gray level co-occurrence matrix (GLCM), which was presented by Haralick [26]. The GLCM is based on the estimation of the second-order statistics. Each entry (i, j) in GLCM corresponds to the number of occurrences of the pair of gray levels i and j separated by distance d and angle θ. The values of d, for which the GLCM is computed, depend on the nature of the texture. Small d values are suitable for fine textures, whereas larger distances are needed to measure coarse textures. Once the distance d is specified, the orientations should be chosen so that there are no symmetrical redundancies (e.g. $\theta = 0$ and $\theta = \pi$ are redundant). If the texture is isotropic, the orientation has even lesser importance, thus, the GLCM's are measured for different angles, and averaged to give a better estimate of the texture.

The co-occurrence matrix is defined as follows: Let I be a 2-D image with N_g gray levels $[1, 2, \ldots, N_g]$ and suppose that X and Y denote the height and width of the image, respectively. As presented before, the GLCMs define the probability of joining the pair of gray levels i and j separated by distance d and angle θ or displacement $\Delta = (\Delta x, \Delta y)$, where $I(x, y) = i$ and $I(x + \Delta x, y + \Delta y) = j$. Usually, the estimated co-occurrence matrix is written as:

$$M(i, j) = \frac{\sum_{(x,y)\in J} \left[\delta(I(x, y) - i)\delta(I(x + \Delta x, y + \Delta y) - j)\right]}{(X - \Delta x) \times (Y - \Delta y)} \tag{1}$$

where δ is the Kronecker symbol and $J = \left\{(x, y) \in I \,|\, (x + \Delta x, y + \Delta y) \in I\right\}$.

3. Band selection based on texture dependent information

Let us consider a set of image bands $A_1, \ldots, A_{N_b}$, where A_k is a random variable representing the image band k, N_b is the number of bands in the multispectral image. Let $M_1, \ldots, M_{N_b}$ be the co-occurrence matrices computed from the image bands $A_1, \ldots, A_{N_b}$ respectively for a given distance d and angle θ. Let x_{ij}^{k} be the (i, j)th element of an co-occurrence matrix M_k, where $1 \le k \le N_b$.

Formally, the mutual information of two co-occurrence matrices M_{k_1} and M_{k_2} can be defined as:

$$I(M_{k_1}; M_{k_2}) = \sum_{x_{ij}^{k_1}, x_{ij}^{k_2}} p(x_{ij}^{k_1}, x_{ij}^{k_2}) \log_2 \frac{p(x_{ij}^{k_1}, x_{ij}^{k_2})}{p_1(x_{ij}^{k_1}) p_2(x_{ij}^{k_2})} \tag{2}$$

where $p(x_{ij}^{k_1}, x_{ij}^{k_2})$ represents a joint probability distribution, $p_1(x_{ij}^{k_1})$ and $p_2(x_{ij}^{k_2})$ are the marginal probability distribution functions of M_{k_1} and M_{k_2} respectively.

The amount of information contained in the two co-occurrence matrices M_{k_1} and M_{k_2} can be expressed as the joint entropy:

$$H(M_{k_1}, M_{k_2}) = \sum_{x_{ij}^{k_1}, x_{ij}^{k_2}} p(x_{ij}^{k_1}, x_{ij}^{k_2}) \log_2 \frac{1}{p(x_{ij}^{k_1}, x_{ij}^{k_2})} \tag{3}$$

The term $\log_2(1/p(x_{ij}^{k_1}, x_{ij}^{k_2}))$ means that the amount of information gained from an event with probability $p(x_{ij}^{k_1}, x_{ij}^{k_2})$ is inversely related to the probability that takes place in this event.

This concept could be generalized in the case of several co-occurrence matrices as follows:

$$H(M_1, \ldots, M_{N_b}) = \sum_{x_{ij}^{1}, \ldots, x_{ij}^{N_b}} p(x_{ij}^{1}, \ldots, x_{ij}^{N_b}) \log_2 \frac{1}{p(x_{ij}^{1}, \ldots, x_{ij}^{N_b})} \tag{4}$$

The rarer is an event, the more meaning is assigned to the occurrence of this event. Therefore, the information per event is weighted by the probability of its occurrence. The resulting entropy term is the average amount of information gained from a set of possible events.

We note that the joint probability distribution can be estimated by the following equation:

$$p(m_1, \ldots, m_{N_b}) = \frac{h(m_1, \ldots, m_{N_b})}{XY} \tag{5}$$

where $h(m_1, \ldots, m_{N_b})$ is the joint gray-level histogram of bands $M_1, \ldots, M_{N_b}$ and the normalizing factor, XY (X columns and Y rows) is the image size, assuming all image bands with equal size.

The proposed algorithm is based on the minimization of the dependent information [20]. The measure of DTI consists of a relation between the joint entropy and the union of the conditional entropies of the considered set of co-occurrence matrices $M_1, \ldots, M_{N_b}$. This criterion can be written mathematically as:

$$\Theta_{\mathrm{DTI}} = H(M_1, \ldots, M_{N_b}) - \sum_{i_1=1}^{n} H(M_{i_1} | M_{i_2}, \ldots, M_{i_n}) \tag{6}$$

where M_{i_1} is the co-occurrence matrix computed from the image band i_1. $M_{i_2}, \ldots, M_{i_n}$ are the complementary matrices of M_{i_1}; $H(M_1, \ldots, M_n)$ is the joint entropy, which represents the total amount of joint information of set of co-occurrence matrices $M_1, \ldots, M_{N_b}$; and $H(M_{i_1} | M_{i_2}, \ldots, M_{i_n})$ is the conditional entropy that represents the amount of independent information in M_{i_1} having measured the rest co-occurrence matrices $M_{i_2}, \ldots, M_{i_n}$, it can be calculated in the following way:

$$H(M_{i_1} | M_{i_2}, \ldots, M_{i_n}) = H(M_{i_1}, M_{i_2}, \ldots, M_{i_n}) - H(M_{i_2}, \ldots, M_{i_n}) \tag{7}$$

In the case of three co-occurrence matrices, the Θ_{DTI} can be represented by entropy Venn diagrams as shown in Fig. 1

4. Proposed band selection strategy

The proposed strategy for selecting optimal band combinations is based on a sequential forward scheme. Given the number of available bands N_b, all possible way combinations are formed by minimizing the DTI given by Eq. (6).

The proposed approach can be summarized as follows:

1. Let $A_1, \ldots, A_{N_b}$ be the initial set of image bands, where N_b is the number of image bands. Let S be the subset of selected bands and N_s be the desired number of selected bands. L_i is the subset of i selected co-occurrence matrices.
2. For a given initial set of image bands $A_1, \ldots, A_{N_b}$, compute the co-occurrence matrix M_i of each band A_i, where $1 \le i \le N_b$. The