Possibilistic Similarity based Image Classification

B. Alsahwa[1,2], S. Almouahed[1], D. Guériot[1,2] and B. Solaiman[1,2]
[1]Image & Information Processing Dept., Telecom Bretagne, Institut Mines-Télécom, Brest, France
[2]Lab-STICC UMR CNRS 3192, Laboratoire en sciences et technologies de l'information, de la communication et de la connaissance, Institut Mines-Télécom-Télécom Bretagne-UE, Brest, France

Keywords: Possibility Theory, Classification, Contextual Information, *a Priori* Knowledge, Possibilistic Similarity.

Abstract: In this study, an approach for image classification based on possibilistic similarity is proposed. This approach, due to the use of possibilistic concepts, enables an important flexibility to integrate both contextual information and a priori knowledge. Possibility distributions are, first, obtained using a priori knowledge given in the form of learning areas delimitated by an expert. These areas serve for the estimation of the probability density functions of different thematic classes. The resulting probability density functions are then transformed into possibility distributions using Dubois-Prade's probability-possibility transformation. Several measures of similarity between classes were tested in order to improve the discrimination between classes. The classification is then performed based on the principle of possibilistic similarity. Synthetic and real images are used in order to evaluate the performances of the proposed model.

1 INTRODUCTION

An accurate and reliable image classification is a crucial task in many applications such as content based image retrieval, medical and remote-sensing image analysis and scene interpretation. Several techniques of image classification are built on a local approach of the scene to deal with, as opposed to those built on segmentation or object (Caloz and Collet, 2001). It is generally accepted that taking into account the geometric dimension, including the context or neighbourhood, contributes to the performance of the classification (Tso and Mather, 2009). Two families of local classification approach can be encountered in the literature: The first family uses a thematic classifier working, first, at pixel-level only, followed by a step of integrating contextual information (Kim, 1996) and (Shaban and Dikshit, 2001). This two-step process constitutes clearly a weakness. Conversely, the other family simultaneously combines the rules of thematic similarity and spatial proximity in a single classification process (Rakotoniaina and Collet, 2010) and (Besag, 1986).

In this paper, pixel-based image classification systems are considered under the closed world assumption. Each pixel from the analyzed image, *I*, is assumed to belong to one, and only one, thematic class from an exhaustive set of M predefined and mutually exclusive classes $\Omega = \{C_1, C_2, ..., C_M\}$. *Prior* knowledge is assumed to be given as a set of learning areas extracted from the considered image and characterizing the M considered classes (from the expert point of view). Using this *prior* knowledge, M class probability density functions are first estimated using the KDE (*Kernel Density Estimation*) approach (Epanechnikov, 1969) and then transformed into M possibility distributions encoding the "expressed" expert knowledge in a possibilistic framework. In the same way, assuming the considered pixel P_0 as being of a "homogeneous sub-area", a local possibility distribution $\pi_{P0}(x)$ will be constructed. This local possibility distribution stands for the possibility degree to observe the pixel P_0 in the considered sub-area. The application of similarity concept on the M possibility distributions will lead, on one hand, to determine the similarity function which maximizes the discrimination between classes, and on the other hand, to enable the classification of sub-areas represented by local possibility distributions

The use of possibilistic concepts increases the capacity as well as the flexibility to deal with uncertainty as, for most real-world problems, the modelled knowledge is affected by different forms of imperfections: imprecision, incompleteness, ambiguity, etc.

Alsahwa B., Almouahed S., Guériot D. and Solaiman B. (2013).
Possibilistic Similarity based Image Classification.
In *Proceedings of the 2nd International Conference on Pattern Recognition Applications and Methods*, pages 271-275
DOI: 10.5220/0004265002710275

In the next section, a brief review of basic concepts of possibility theory is introduced. Study of different similarity functions to quantify the similarity between classes is the subject of the third section. The proposed approach will be detailed in the forth section. Section 5 is devoted to the experimental results obtained when the proposed approach is applied using synthetic as well as real images.

2 POSSIBILITY THEORY

Possibility theory was first introduced by Zadeh in 1978 as an extension of fuzzy sets and fuzzy logic theory to express the intrinsic fuzziness of natural languages as well as uncertain information (Zadeh, 1978). In the case where the available knowledge is ambiguous and encoded as a membership function into a fuzzy set defined over the decision set, the possibility theory transforms each membership value into a possibilistic interval of possibility and necessity measures (Dubois and Prade, 1980).

2.1 Possibility Distribution

Let us consider an exclusive and exhaustive universe of discourse $\Omega = \{C_1, C_2, ..., C_M\}$ formed by M elements C_m, m = 1, ..., M (e.g., thematic classes, hypothesis, elementary decisions, etc). Exclusiveness means that one and only one element may occur at time, whereas, exhaustiveness refers to the fact that the occurring element certainly belongs to Ω. A key feature of possibility theory is the concept of a *possibility distribution*, denoted by π, assigning to each element $C_m \in \Omega$ a value from a bounded set [0,1] (or a set of graded values). This value $\pi(C_m)$ encodes our state of knowledge, or belief, about the real world and represents the possibility degree for C_m to be the unique occurring element.

2.2 Possibility Distributions Estimation based on Pr-π Transformation

Two approaches are generally used for the estimation of a possibility distribution. The first approach consists on using standard forms predefined in the framework of fuzzy set theory for membership functions (i.e. triangular, Gaussian, trapezoidal, etc.), and tuning the form parameters using a manual or an automatic tuning method. The second possibility distributions estimation approach is based on the use of statistical data where an uncertainty function (e.g. histogram, probability distribution function, basic belief function, etc.); is first estimated and then transformed into a possibility distribution

As we consider that the available expert's knowledge is expressed through the definition of learning areas representing different thematic classes, i.e. statistical data, we will focus on the second estimation approach. Several Pr-π transformations are proposed in the literature. Dubois *et al.* (Dubois and Prade, 1983) suggested that any Pr-π transformation of a probability distribution function, Pr, into a possibility distribution, π, should be guided by the two following principles:

- The probability-possibility *consistency principle*:

$$\Pi(A) \geq \Pr(A), \quad \forall A \subseteq \Omega \tag{1}$$

- The *preference preservation principle*:

$$\Pr(A) < \Pr(B) \Leftrightarrow \Pi(A) < \Pi(B), \quad \forall A, B \subseteq \Omega \tag{2}$$

Verifying these two principles, a Pr-π transformation has been suggested by Dubois *et al.* (Dubois and Prade, 1983):

$$\pi(C_m) = \Pi(\{C_m\}) = \sum_{j=1}^{M} \min\left[\Pr(\{C_j\}),\ \Pr(\{C_m\})\right] \tag{3}$$

In our study, this transformation is considered in order to transform the probability distributions into possibility distributions.

3 SIMILARITY MEASURES

The issue of comparing imperfect pieces of information depends on the way these pieces of information are represented. In the case of possibility theory, comparing uncertain pieces of information comes down to comparing possibility distributions representing these pieces of information.

Considering the expert's predefined set of M thematic classes contained in the analyzed image, $\Omega = \{C_1, C_2 ..., C_M\}$, a set of M possibility distributions can be defined as follows:

$$\pi_{C_m} : D \to [0,1]$$
$$x(\mathrm{P}) \to \pi_{C_m}(x(\mathrm{P}))$$

where D refers to the definition domain of the observed feature x(P). For each class C_m, $\pi_{C_m}(x(\mathrm{P}))$ associates each pixel $\mathrm{P} \in I$, observed through a feature $x(P) \in D$, with a possibility degree of belonging to the class C_m, m = 1, ..., M.

Considering two classes C_m and C_n of the set Ω, different possibilistic similarity and distance functions "Sim" can be defined between their two possibility distributions π_{Cm} and π_{Cn}. The behaviour of these functions can be studied in order to obtain a better discrimination between classes C_m and C_n. To do this, calculating a similarity matrix $\mathrm{Sim}(\pi_{Cm}, \pi_{Cn})$ informs us about such inter-classes behaviour and will help in choosing the right measure in the given context:

$$S=\begin{pmatrix} \mathrm{Sim}\left(\pi_{C_m},\pi_{C_m}\right) & \mathrm{Sim}\left(\pi_{C_m},\pi_{C_n}\right) \\ \mathrm{Sim}\left(\pi_{C_n},\pi_{C_m}\right) & \mathrm{Sim}\left(\pi_{C_n},\pi_{C_n}\right) \end{pmatrix} \quad (4)$$

3.1 Possibilistic Similarity Functions

This subsection is devoted to review some existing possibilistic similarity and distance functions which are the most frequently encountered in the literature:

- *Information closeness*: this similarity measure was proposed by (Higashi and Klir, 1983):

$$G(\pi_{C_m},\pi_{C_n})=g(\pi_{C_m},\pi_{C_m} \vee \pi_{C_n}) + g(\pi_{C_n},\pi_{C_m} \vee \pi_{C_n}) \quad (5)$$

where $g(\pi_{C_m}, \pi_{C_n})=U(\pi_{C_n})-U(\pi_{C_m})$. $\vee$ is taken as maximum operator and U is the non-specificity measure. Given an ordered possibility distribution π such that $1= \pi_1 \geq \pi_2 \geq \ldots \geq \pi_n$ the U of π is given by:

$$U(\pi)=[\sum_{i=1}^{n}(\pi_i - \pi_{i+1})\log_2 i]+(1-\pi_1)\log_2 n \quad (6)$$

where $\pi_{n+1}=0$ by convention. Hence the similarity measure based on the *Information closeness* is given by:

$$\mathrm{Sim}_G\left(\pi_{C_m},\pi_{C_n}\right) = 1 - \frac{G(\pi_{C_m},\pi_{C_n})}{G_{\max}} \quad (7)$$

- *Minkowski distance*: Since possibility distributions are often represented as vectors, the most popular metrics for possibility distributions are induced by the *Minkowski* norm (Lp) which is used in vector spaces.

$$L_p\left(\pi_{C_m},\pi_{C_n}\right)=\sqrt[p]{\sum_{i=1}^{|D|}\left|\pi_{C_m}\left(x_i\right)-\pi_{C_n}\left(x_i\right)\right|^p} \quad (8)$$

Three particular cases of equation (10) are often investigated: L_1-norm (*Manhattan distance*), L_2-norm (*Euclidean distance*), and L_∞-norm (*Maximum distance*). These cases of *Minkowski distance* can be transformed into similarity measure by the following:

$$\mathrm{Sim}_p\left(\pi_{C_m},\pi_{C_n}\right) = 1 - \frac{L_p}{\sqrt[p]{|D|}} \quad (9)$$

- *Information affinity*: this similarity measure was proposed by Jenhani *et al.* (Jenhani *et al.*, 2007)

$$\mathrm{Sim}_{IA}\left(\pi_{C_m},\pi_{C_n}\right) = 1 - \frac{\kappa \times L_p(\pi_{C_m},\pi_{C_n}) + \lambda \times \mathrm{Inc}(\pi_{C_m},\pi_{C_n})}{\kappa+\lambda} \quad (10)$$

Where $\kappa>0$ and $\lambda>0$, $\mathrm{Inc}(\pi_{C_m}, \pi_{C_n})$ represents the inconsistency degree between π_{C_m} and π_{C_n} defined as follows

$$\mathrm{Inc}(\pi_{C_m},\pi_{C_n}) = 1 - \max(\min(\pi_{C_m}, \pi_{C_n}) \quad (11)$$

3.2 Evaluation of the Similarity between Two Classes

A 100×100 synthetic image composed of two thematic classes is generated in order to evaluate the similarity between two classes. The intensity of the pixels from C_1 and C_2 are generated as two Gaussian distributions $\mathbb{G}(m_1=110, \sigma_1=10)$ and $\mathbb{G}(m_2=120, \sigma_2=10)$(Figure 1)

The evaluation principle of the similarity between the two classes is to retain the similarity function whose similarity matrix is the closest to the identity matrix I_2 in term of Euclidean distance D:

$$D=\sqrt{\sum_{i,j}\left[S(i,j)-I_2(i,j)\right]^2} \quad (12)$$

D was calculated for each similarity function by firstly varying the mean of the class C_2 and then the standard deviation of class C_2, while maintaining a fixed value for the mean and standard deviation of the class C_1(Figure 1).

From the curves in Figure 1, the similarity function called "Maximum distance" $\mathrm{Sim}_\infty(\pi_{C_m}, \pi_{C_n})$ tends to the identity matrix faster than the other functions when the studied values m_2-m_1 and σ_2-σ_1 increase.

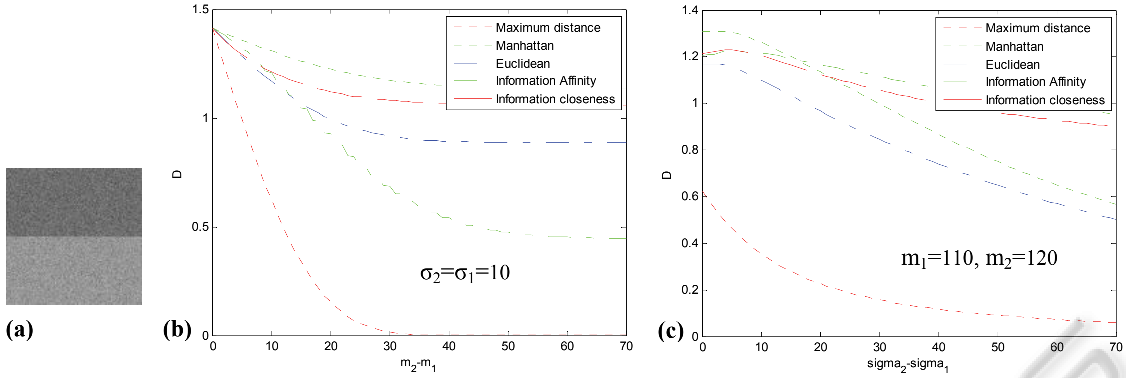

Figure 1: (a) Synthetic image (b) Evolution of the measure D as a function of the mean difference between classes C_1 and C_2 (c) Evolution of the measure D as a function of the difference of deviations between classes C_1 and C_2.

4 THE PROPOSED CLASSIFICATION APPROACH

As previously detailed, the samples initial set, considered by the expert, is used in order to estimate the probability density functions of different thematic classes, which in turns are transformed into possibility distributions through the application of the Pr-π Dubois-Prade's transformation.

The estimation of these M possibility distributions forms the first step in the proposed approach. The second step consists in the classification of each pixel of the analyzed image *I* by firstly estimating the local possibility distribution around the pixel of interest P_0. Second, the process of assigning a class to the considered pixel P_0 is to determine the nearest class via the similarity function Sim_∞ used to measure the similarity between this pixel's local possibility distribution and possibility distributions of each of the M classes (Figure 2).

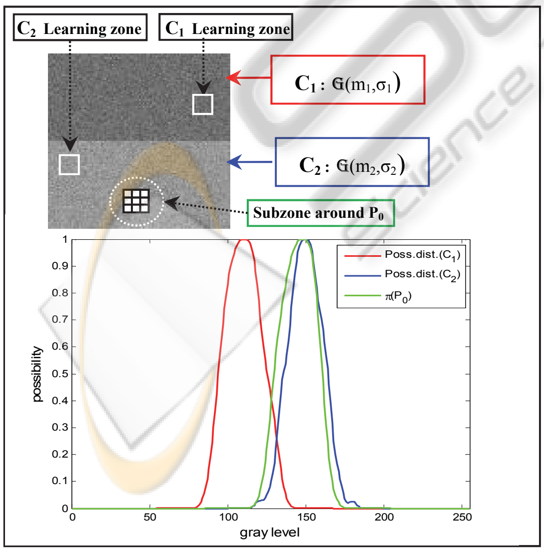

Figure 2: Synthetic image, possibility distributions of classes C_1, C_2 and the local possibility distribution in a subzone around the pixel of interest P_0.

5 EXPERIMENTAL RESULTS

5.1 Simulated Data

For the experimental evaluation purpose, a new synthetic image of size 96×128 pixel is generated (Figure 3). Pixels from C_1 and C_2 are generated as two Gaussian distributions $\mathfrak{G}(m_1=125,\ \sigma_1=15)$ and $\mathfrak{G}(m_2=100,\ \sigma_2=20)$. This synthetic image is classified using the proposed approach and the Bayesian approach (Hand, 1981), respectively. The

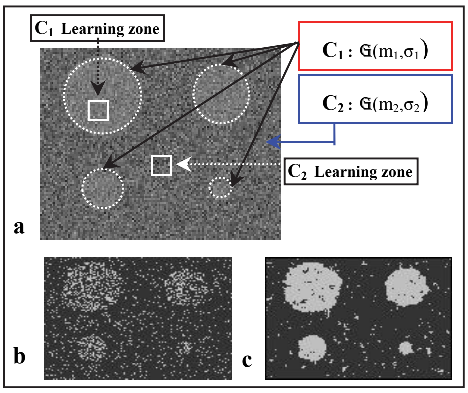

Figure 3: (a) Synthetic image (b) Bayesian classification (c) Proposed approach classification.

classification error rate when using the possibilistic approach with Sim_∞ function is 8.5% while the error rate obtained by the Bayesian approach is 18.3%.

5.2 Medical Application

The proposed approach of classification is applied on a mammographic image composed of two classes: tumor and normal tissue (figure 4). This image is extracted from the MIAS image database (Mammographic Image Analysis Society).

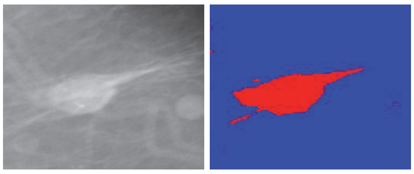

Figure 4: (left) A mammographic image composed of two classes, (right) Classified image using the proposed approach.

A visual analysis of the obtained results shows that the proposed approach allows obtaining an interesting homogeneity of the regions determined from samples based on measures limited to windows of size 3 \times 3.

6 CONCLUSIONS

In this study, a classification approach was developed based on the possibility theory that enables the integration of contextual information and a *priori* knowledge. Indeed, one of the key points of the proposed approach is to characterize the pixel to be classified taking into account its neighbourhood through the creation of local possibility distribution. Another key point of our approach is to propose a classification method based on the similarity between class possibility distribution and local possibility distribution, and not on a membership degree, of parameters extracted from the local window, to possibility distributions of classes. The first results on both the synthetic image and the real medical image (compared to the results obtained using a Bayesian approach) seem promising.

REFERENCES

Besag, J., 1986. On the statistical analysis of dirty pictures. *Journal of the Royal Society*, series B, vol. 48, pp. 259-302.

Caloz, R., and Collet, C., 2001. Précis de télédétection : Traitements numériques d'images de télédétection. *Presses de l'Université du Québec*, Vol.3, 386 p, Canada.

Dubois, D., Prade, H., 1980.Fuzzy Sets and Systems: Theory and Applications. *Academic Press*, New York.

Dubois, D., Prade, H., 1983. Unfair Coins and Necessity Measures: towards a possibilistic Interpretation of Histograms. *Fuzzy Sets and Syst*. Vol.10, pp. 15-20.

Epanechnikov, V. A., 1969. Non-parametric estimation of a multivariate probability density. *Theory of Probability and its Applications* 14: 153–158.

Higashi, M., Klir, G. J., 1983. Measures of uncertainty and information based on possibility distributions. *International Journal of General Systems*, 9 (1), 43-58.

Hand, D. J., 1981. Discrimination and Classification. *Wiley Series in Probability and Mathematical Statistics*.

Jenhani, I., Ben Amor, N., Elouedi, Z., Benferhat, S., Mellouli, K., 2007. Information affinity: A new similarity measure for possibilistic uncertain information. *In Proceedings of the 9th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, 840-852.

Kim, K. E., 1996. Adaptive majority filtering for contextual classification of remote sensing data. *international journal of remote sensing*, Vol. 17, pp. 1083-1087.

Rakotoniaina, S.,and Collet, C., 2010. Amélioration de la Qualité de la Classification d'une Image Multispectrale à l'aide d'un classificateur contextuel, *Revue Télédétection*, Vol. 9, pp. 259-27.

Shaban, M.A., and Dikshit, O.,2001. Improvement of classification in urban areas by the use of textural features : the case of Lucknow city, Uttar Pradesh, *International Journal of Remote Sensing*, vol. 22, pp. 565-593.

Tso, B., and Mather, P. M., 2009. *classification methods for remotely sensed data*. taylor & francis group.

Zadeh, L. A., 1978. Fuzzy Sets as a Basis for a Theory of possibility. *Fuzzy Sets Syst*., vol. 1, PP.3-28, 1978.