Feature Discovery through Hierarchies of Rough Fuzzy Sets

Alfredo Petrosino and Alessio Ferone

Abstract. Rough set theory and fuzzy logic are mathematical frameworks for granular computing forming a theoretical basis for the treatment of uncertainty in many real–world problems, including image and video analysis. The focus of rough set theory is on the ambiguity caused by limited discernibility of objects in the domain of discourse; granules are formed as objects and are drawn together by the limited discernibility among them. On the other hand, membership functions of fuzzy sets enables efficient handling of overlapping classes. The hybrid notion of rough fuzzy sets comes from the combination of these two models of uncertainty and helps to exploit, at the same time, properties like coarseness, by invoking rough sets, and vagueness, by considering fuzzy sets. We describe a model of the hybridization of rough and fuzzy sets, that allows for further refinements of rough fuzzy sets. This model offers viable and effective solutions to some problems in image analysis, e.g. image compression.

Keywords: Rough Fuzzy Sets, Modelling Hierarchies, Feature Discovery, Image Analysis.

1 Introduction

Granular computing is based on the concept of information granule, that is a collection of similar objects which can be considered as indistinguishable. Partition of an universe into granules offers a coarse view of the universe where concepts, represented as subsets, can be approximated by means of granules. In this framework, rough set theory can be regarded to as a family

Alfredo Petrosino · Alessio Ferone
Department of Applied Science
University of Naples "Parthenope"
Centro Direzionale Isola C4
80143 Naples, Italy
e-mail: {alfredo.petrosino,alessio.ferone}@uniparthenope.it

W. Pedrycz and S.-M. Chen (Eds.): Granular Computing and Intell. Sys., ISRL 13, pp. 57–73.
springerlink.com

of methodologies and techniques that make use of granules [9, 10]. The focus of rough set theory is on the ambiguity caused by limited discernibility of objects in the domain of discourse. Granules are formed as objects and are drawn together by the limited discernibility among them. Granulation is of particular interest when a problem involves incomplete, uncertain or vague information. In such cases, precise solutions can be difficult to obtain and hence the use of techniques based on granules can lead to a simplification of the problem at hand.

At the same time, multivalued logic can be applied to handle uncertainty and vagueness present in information systems, the most visible of which is the theory of fuzzy sets [13]. In this framework, uncertainty is modelled by means of functions that define the degree of belonginess of an object to a given concept. Hence membership functions of fuzzy sets enable efficient handling of overlapping classes.

The hybrid notion of rough fuzzy sets comes from the combination of these two models of uncertainty to exploit, at the same time, properties like coarseness, by handling rough sets [9], and vagueness, by handling fuzzy sets [13]. In this combined framework, rough sets embody the idea of indiscernibility between objects in a set, while fuzzy sets model the ill-definition of the boundary of a subclass of this set. Combining both notions leads to consider, as instance, approximation of sets by means of similarity relations or fuzzy partitions. The rough fuzzy synergy is hence adopted to better represent the uncertainty in granular computation.

Nevertheless, some considerations are in order. Classical rough set theory is defined over a given partition, although several equivalence relations, and hence partitions, can be defined over the universe of discourse. Different partitions correspond to a coarser or finer view of the universe, because of different information granules, thus leading to coarser or finer definition of the concept to be provided. Then a substantial interest arises about the possibility of exploiting different partitions and, possibly, rough sets of higher order. Some approaches have been presented to exploit hierarchical granulation [5] where various approximations are obtained with respect to different levels of granulation. Considered as a nested sequence of granulations by a nested sequence of equivalence relations, this procedure leads to a nested sequence of rough set approximations and to a more general approximation structure. Hierarchical representation of the knowledge is also used in [7] to build a sequence of finer reducts so to obtain multiple granularities at multiple layers. The hierarchical reduction can handle problem with coarser granularity at lower level so to avoid incompleteness of data present in finer granularity at deeper layer. A different approach is presented in [8] where authors report a Multi-Granulation model of Rough-Set (MGRS) as an extension of Pawlak's rough set model. Moreover, this new model is used to define the concept of approximation reduct as the smallest attribute subset that preserves the lower approximation and upper approximation of all decision classes in MGRS.

The hybridization of rough and fuzzy sets reported here has been observed to possess a viable and effective solution to some of the most difficult problems in image analysis. The model exhibits a certain advantage of having a new operator to compose rough fuzzy sets, called $\mathcal{RF}$-product, able to produce a sequence of composition of rough fuzzy sets in a hierarchical manner. Theoretical foundations and properties, together with an example of application for image compression are described in the following sections.

2 Rough Fuzzy Sets: A Background

Let us start from the definition of a rough fuzzy set given by Dubois and Prade [6]. Let U be the universe of discourse, X a fuzzy subset of U, such that $\mu_X(u)$ represents the fuzzy membership function of X over U, and R an equivalence relation that induces the partition $U/R = \{Y_1, \ldots, Y_p\}$ (from now on denoted as $\mathcal{Y}$) over U in p disjoint sets, i.e. $Y_i \bigcap Y_j = \emptyset \; \forall i, j = 1, \ldots, p$ and $\bigcup_{i=1}^{p} Y_i = U$. The lower and upper approximation of X by R, i.e. $\underline{R}(X)$ and $\overline{R}(X)$ respectively, are fuzzy sets defined as

$$\mu_{\underline{R}(X)}(Y_i) = \inf\{\mu_X(u) | Y_i = [u]_R\} \tag{1}$$

$$\mu_{\overline{R}(X)}(Y_i) = \sup\{\mu_X(u) | Y_i = [u]_R\} \tag{2}$$

i.e. $[u]_R$ is a set such that (1) and (2) represent the degrees of membership of Y_i in $\underline{R}(X)$ and $\overline{R}(X)$, respectively. The couple of sets $< \underline{R}(X), \overline{R}(X) >$ is called *rough-fuzzy set* denoting a fuzzy concept (X) defined in a crisp approximation space (U/R) by means of two fuzzy sets ($\underline{R}(X)$ and $\overline{R}(X)$). Specifically, identifying $\pi_i(u)$ as the function that returns 1 if $u \in Y_i$ and 0 if $u \notin Y_i$, and considering $Y_i = [u]_R$ and $\pi_i(u) = 1$, the following relationships hold:

$$\mu_{\underline{R}(X)}(Y_i) = \inf_u max\{1 - \pi_i(u), \mu_X(u)\} \tag{3}$$

$$\mu_{\overline{R}(X)}(Y_i) = \sup_u min\{\pi_i(u), \mu_X(u)\} \tag{4}$$

To emphasize that the lower and upper approximations of the fuzzy subset X are, respectively, the infimum and the supremum of the membership functions of the elements of a class Y_i to the fuzzy set X, we can define a rough-fuzzy set as a triple

$$RF_X = (\mathcal{Y}, \mathcal{I}, \mathcal{S}) \tag{5}$$

where $\mathcal{Y} = \{Y_1, \ldots, Y_p\}$ is a partition of U in p disjoint subsets $Y_1, \ldots, Y_p$, and $\mathcal{I}, \mathcal{S}$ are mappings of kind $U \to [0, 1]$ such that $\forall u \in U$,

$$\mathcal{I}(u) = \sum_{i=1}^{p} \underline{\nu_i} \times \mu_{Y_i}(u) \tag{6}$$

$$\mathcal{S}(u) = \sum_{i=1}^{p} \overline{\nu_i} \times \mu_{Y_i}(u) \tag{7}$$

where

$$\underline{\nu}_i = \inf\{\mu_X(u)|u \in Y_i\} \tag{8}$$

$$\overline{\nu}_i = \sup\{\mu_X(u)|u \in Y_i\} \tag{9}$$

for the given subsets $\mathcal{Y} = \{Y_1, \ldots, Y_p\}$ and for every choice of function $\mu : U \to [0, 1]$. $\mathcal{Y}$ and μ uniquely define a rough-fuzzy set as stated below

Definition 1. Given a subset $X \subseteq U$, if μ is the membership function μ_X defined on X and the partition $\mathcal{Y}$ is made with respect to an equivalence relation $\mathcal{R}$, i.e. $\mathcal{Y} = U/\mathcal{R}$, then X is a fuzzy set with two approximations $\overline{R}(X)$ and $\underline{R}(X)$, which are again fuzzy sets with membership functions defined as (8) and (9), i.e. $\underline{\nu}_i = \mu_{\underline{R}(X)}$ and $\overline{\nu}_i = \mu_{\overline{R}(X)}$. The pair of sets $< \overline{R}(X), \underline{R}(X) >$ is then a *rough fuzzy set*.

Let us recall the generalized definition of rough set given in [1]. Expressions for the lower and upper approximations of a given set X are

$$\underline{R}(X) = \{(u, \mathcal{I}(u))|u \in U\} \tag{10}$$

$$\overline{R}(X) = \{(u, \mathcal{S}(u))|u \in U\} \tag{11}$$

$\mathcal{I}$ and $\mathcal{S}$ are defined as:

$$\mathcal{I}(u) = \sum_{i=1}^{p} \mu_{Y_i}(u) \times \inf_{\varphi \in U} \max(1 - \mu_{Y_i}(\varphi), \mu_X(\varphi)) \tag{12}$$

$$\mathcal{S}(u) = \sum_{i=1}^{p} \mu_{Y_i}(u) \times \sup_{\varphi \in U} \min(\mu_{Y_i}(\varphi), \mu_X(\varphi)) \tag{13}$$

where μ_{Y_i} is the membership degree of each element $u \in U$ to a granule $Y_i \in U/R$ and μ_X is the membership function associated with X.

If we rewrite (8) and (9) as

$$\underline{\nu}_i = \mu_{\underline{R}(X)}(Y_i) = \inf_{\varphi \in U} \max(1 - \mu_{Y_i}(\varphi), \mu_X(\varphi)) \tag{14}$$

$$\overline{\nu}_i = \mu_{\overline{R}(X)}(Y_i) = \sup_{\varphi \in U} \min(\mu_{Y_i}(\varphi), \mu_X(\varphi)) \tag{15}$$

and considering a Boolean equivalence relation R, we arrive at the same definition of rough fuzzy set as given in (3) and (4). Indeed, considering (14) and the equivalence relation R, $\mu_Y(\varphi)$ takes values in $\{0, 1\}$ hence the expression $1 - \mu_Y(\varphi)$ equals 0 if $\varphi \in Y$ or 1 if $\varphi \notin Y$. Furthermore the max operation returns 1 or $\mu_X(\varphi)$ depending on the fact that $\varphi \in Y$ or $\varphi \notin Y$. The operation inf then returns the infimum of such values, that is the minimum value of $\mu_X(\varphi)$. The same applies to (15).

3 Hierarchical Refinement of Rough-Fuzzy Sets

Rough set theory allows to partition the given data into equivalence classes. Nevertheless, given a set U, it is possible to employ different equivalence relations and hence produce different data partitions. This leads to a choice of the partition that represents the data in the best manner. For example, let us consider N-dimensional patterns, with $N = 4$ as in Table 1.

Table 1 Example of data

	A_1	A_2	A_3	A_4
u_1	a	a	b	c
u_2	b	c	c	c
u_3	a	b	c	a
u_4	c	b	a	b

Let $\mathcal{Y}_i$ be the partition obtained applying the equivalence relation R_{A_i} on the attribute A_i. We may get from Table 1 the following four partitions

$$\begin{aligned}
\mathcal{Y}^1 &= \{\{u_1, u_3\}, \{u_2\}, \{u_4\}\} \\
\mathcal{Y}^2 &= \{\{u_1\}, \{u_2\}, \{u_3, u_4\}\} \\
\mathcal{Y}^3 &= \{\{u_1\}, \{u_2, u_3\}, \{u_4\}\} \\
\mathcal{Y}^4 &= \{\{u_1, u_2\}, \{u_3\}, \{u_4\}\}
\end{aligned} \tag{16}$$

that, without any apriori knowledge, have potentially the same data representation power. To exploit all the possible partitions by means of simple operations, we propose to refine them in a hierarchical manner, so that partitions at each level of the hierarchy retain all the important informations contained into the partitions of the lower levels. The operation employed to perform the hierarchical refinement is called *Rough–Fuzzy product* ($\mathcal{RF}$-product) and is defined by:

Definition 2. Let $RF^i = (\mathcal{Y}^i, \mathcal{I}^i, \mathcal{S}^i)$ and $RF^j = (\mathcal{Y}^j, \mathcal{I}^j, \mathcal{S}^j)$ be two rough fuzzy sets defined, respectively, over partitions $\mathcal{Y}^i = (Y_1^i, \dots, Y_p^i)$ and $\mathcal{Y}^j = (Y_1^j, \dots, Y_p^j)$ with $\mathcal{I}^i$ (resp. $\mathcal{I}^j$) and $\mathcal{S}^i$ (resp. $\mathcal{S}^j$) indicating the measures expressed in Eqs. (6) and (7). The *$\mathcal{RF}$–product* between two rough-fuzzy sets, denoted by $\otimes$, is defined as a new rough fuzzy set

$$RF^{i,j} = RF^i \otimes RF^j = (\mathcal{Y}^{i,j}, \mathcal{I}^{i,j}, \mathcal{S}^{i,j})$$

where $\mathcal{Y}^{i,j} = (Y_1^{i,j}, \ldots, Y_{2p-1}^{i,j})$ is a new partition whose equivalence classes are

$$Y_k^{ij} = \begin{cases} \bigcup_{\substack{s=1\\q=h}}^{\substack{s=h\\q=1}} Y_q^i \cap Y_s^j & h = k, \quad k \leq p \\ \\ \bigcup_{\substack{s=h\\q=p}}^{\substack{s=p\\q=h}} Y_q^i \cap Y_s^j & h = k - p + 1, \quad k > p \end{cases} \tag{17}$$

and $\mathcal{I}^{i,j}$ and $\mathcal{S}^{i,j}$ are

$$\mathcal{I}^{i,j}(u) = \sum_{k=1}^{2p-1} \underline{\nu}_k^{i,j} \times \mu_k^{i,j}(u) \tag{18}$$

$$\mathcal{S}^{i,j}(u) = \sum_{k=1}^{2p-1} \overline{\nu}_k^{i,j} \times \mu_k^{i,j}(u) \tag{19}$$

and

$$\underline{\nu}_k^{ij} = \begin{cases} \sup_{\substack{s=1,\ldots,h\\q=h,\ldots,1}} \{\underline{\nu}_q^i, \underline{\nu}_s^j\} & h = k, \quad k \leq p \\ \\ \sup_{\substack{s=h,\ldots,p\\q=p,\ldots,h}} \{\underline{\nu}_q^i, \underline{\nu}_s^j\} & h = k - p + 1, \quad k > p \end{cases} \tag{20}$$

$$\overline{\nu}_k^{ij} = \begin{cases} \inf_{\substack{s=1,\ldots,h\\q=h,\ldots,1}} \{\overline{\nu}_q^i, \overline{\nu}_s^j\} & h = k, \quad k \leq p \\ \\ \inf_{\substack{s=h,\ldots,p\\q=p,\ldots,h}} \{\overline{\nu}_q^i, \overline{\nu}_s^j\} & h = k - p + 1, \quad k > p \end{cases} \tag{21}$$

Let us pick up the example shown in Table 1, and consider partitions $\mathcal{Y}^1$ and $\mathcal{Y}^2$ obtained from equivalence relations R_{A_1} and R_{A_2} defined on U by attributes A_1 and A_2, respectively. In terms of rough–fuzzy sets they are $RF^1 = (\mathcal{Y}^1, \mathcal{I}^1, \mathcal{S}^1)$ and $RF^2 = (\mathcal{Y}^2, \mathcal{I}^2, \mathcal{S}^2)$. Partitions $\mathcal{Y}^1$ and $\mathcal{Y}^2$ are defined as follows

$$\{u_4\} = Y_1^1 \qquad \{u_3, u_4\} = Y_1^2$$
$$\{u_2\} = Y_2^1 \qquad \{u_2\} = Y_2^2$$
$$\{u_1, u_3\} = Y_3^1 \qquad \{u_1\} = Y_3^2$$

The refined partition $\mathcal{Y}^{1,2}$ defined on U by both attributes, corresponds to the partition obtained by $\mathcal{RF}$-producting RF^1 and RF^2.

The new partition $\mathcal{Y}^{1,2}$ is obtained by (in matrix notation)

–	–	$(Y_3^1 \cap Y_1^2)$	$(Y_2^1 \cap Y_1^2)$	$(Y_1^1 \cap Y_1^2)$
–	$(Y_3^1 \cap Y_2^2)$	$(Y_2^1 \cap Y_2^2)$	$(Y_1^1 \cap Y_2^2)$	–
$(Y_3^1 \cap Y_3^2)$	$(Y_2^1 \cap Y_3^2)$	$(Y_1^1 \cap Y_3^2)$	–	–

The final partition is obtained by joining sets by column as explained in Eq. 17

$$\begin{aligned}
Y_1^{1,2} &= \{(Y_1^1 \cap Y_1^2)\} \\
Y_2^{1,2} &= \{(Y_2^1 \cap Y_1^2) \cup (Y_1^1 \cap Y_2^2)\} \\
Y_3^{1,2} &= \{(Y_3^1 \cap Y_1^2) \cup (Y_2^1 \cap Y_2^2) \cup (Y_1^1 \cap Y_3^2)\} \\
Y_4^{1,2} &= \{(Y_3^1 \cap Y_2^2) \cup (Y_2^1 \cap Y_3^2)\} \\
Y_5^{1,2} &= \{(Y_3^1 \cap Y_3^2)\}
\end{aligned}$$

Hence

$$\mathcal{Y}^{1,2} = \{Y_1^{1,2}, Y_2^{1,2}, Y_3^{1,2}, Y_4^{1,2}, Y_5^{1,2}\}$$

and $\mathcal{I}$ and $\mathcal{S}$ of the new rough–fuzzy set, computed as in (18) and (19), are

$$\begin{aligned}
\underline{\nu}_1^{1,2} &= \sup\{\inf\{\underline{\nu}_1^1, \underline{\nu}_1^2\}\} \\
\overline{\nu}_1^{1,2} &= \inf\{\sup\{\overline{\nu}_1^1, \overline{\nu}_1^2\}\} \\
\underline{\nu}_2^{1,2} &= \sup\{\inf\{\underline{\nu}_2^1, \underline{\nu}_1^2\}, \inf\{\underline{\nu}_1^1, \underline{\nu}_2^2\}\} \\
\overline{\nu}_2^{1,2} &= \inf\{\sup\{\overline{\nu}_2^1, \overline{\nu}_1^2\}, \sup\{\overline{\nu}_1^1, \overline{\nu}_2^2\}\} \\
\underline{\nu}_3^{1,2} &= \sup\{\inf\{\underline{\nu}_3^1, \underline{\nu}_1^2\}, \inf\{\underline{\nu}_2^1, \underline{\nu}_2^2\}, \inf\{\underline{\nu}_1^1, \underline{\nu}_3^2\}\} \\
\overline{\nu}_3^{1,2} &= \inf\{\sup\{\overline{\nu}_3^1, \overline{\nu}_1^2\}, \sup\{\overline{\nu}_2^1, \overline{\nu}_2^2\}, \sup\{\overline{\nu}_1^1, \overline{\nu}_3^2\}\} \\
\underline{\nu}_4^{1,2} &= \sup\{\inf\{\underline{\nu}_3^1, \underline{\nu}_2^2\}, \inf\{\underline{\nu}_2^1, \underline{\nu}_3^2\}\} \\
\overline{\nu}_4^{1,2} &= \inf\{\sup\{\overline{\nu}_3^1, \overline{\nu}_2^2\}, \sup\{\overline{\nu}_2^1, \overline{\nu}_3^2\}\} \\
\underline{\nu}_5^{1,2} &= \sup\{\inf\{\underline{\nu}_3^1, \underline{\nu}_3^2\}\} \\
\overline{\nu}_5^{1,2} &= \inf\{\sup\{\overline{\nu}_3^1, \overline{\nu}_3^2\}\}
\end{aligned} \tag{22}$$

The rough–fuzzy set obtained by $RF^1 \otimes RF^2$ is thus defined by

$$RF^{1,2} = (\mathcal{Y}^{1,2}, \mathcal{I}^{1,2}, \mathcal{S}^{1,2})$$

where

$$\mathcal{I}^{1,2}(u) = \sum_{i=1}^{5} \underline{\nu}_i^{1,2} \times \mu_{Y_i^{1,2}}(u)$$
$$\mathcal{S}^{1,2}(u) = \sum_{i=1}^{5} \overline{\nu}_i^{1,2} \times \mu_{Y_i^{1,2}}(u) \tag{23}$$

In case of partitions of different sizes, it is sufficient to add empty sets to have partitions of the same size.

4 Characterization of $\mathcal{RF}$–product

Let us recall that a partition $\mathcal{Y}$ of a finite set U is a collection $\{Y_1, Y_2, \ldots, Y_p\}$ of nonempty subsets (equivalence classes) such that

$$Y_i \cap Y_j = \emptyset \qquad \forall i, j = 1, \ldots, p \tag{24}$$
$$\bigcup_i Y_i = U \tag{25}$$

Hence, each partition defines an equivalence relation and, conversely, an equivalence relation defines a partition, such that the classes of the partition correspond to the equivalence classes of the relation.

Partitions are partially ordered by reverse refinement $\mathcal{Y}^i \subseteq \mathcal{Y}^j$. We say that $\mathcal{Y}^i$ is finer than $\mathcal{Y}^j$ if every equivalence class of $\mathcal{Y}^i$ is contained in some equivalence class of $\mathcal{Y}^j$, that is, for each equivalence class Y_h^j of $\mathcal{Y}^j$, there are equivalence classes $Y_1^i, \ldots, Y_p^i$ of $\mathcal{Y}^i$ such that $Y_h^j = Y_1^i, \ldots, Y_p^i$. If $E(\mathcal{Y}^i)$ is the equivalence relation defined by the partition $\mathcal{Y}^i$, then $\mathcal{Y}^i \subseteq \mathcal{Y}^j$ iff $\forall u, u' \in U, (u, u') \in E(\mathcal{Y}^i) \implies (u, u') \in E(\mathcal{Y}^j)$, that is, $E(\mathcal{Y}^i) \subseteq E(\mathcal{Y}^j)$.
The set $\Pi(U)$ of partitions of a set U forms a lattice under the partial order of reverse refinement. The minimum is the partition where an equivalence relation is a singleton, while the maximum is the partition composed by one single equivalence relation. The meet $\mathcal{Y}^i \wedge \mathcal{Y}^j \in \Pi(U)$ is the partition whose equivalence classes are given by $Y_k^i \cap Y_h^j \neq \emptyset$, where Y_k^i and Y_h^j are equivalence classes of $\mathcal{Y}^i$ and $\mathcal{Y}^j$, respectively. In terms of equivalence relations

$$R_{\mathcal{Y}^i \wedge \mathcal{Y}^j} = R_{\mathcal{Y}^i} \cap R_{\mathcal{Y}^j} \tag{26}$$

is the largest equivalence relation contained in both $R_{\mathcal{Y}^i}$ and $R_{\mathcal{Y}^j}$. The join $\mathcal{Y}^i \vee \mathcal{Y}^j$ is a partition composed by the equivalence classes of the transitive closure of the union of the equivalence relations defined by $\mathcal{Y}^i$ and $\mathcal{Y}^j$. In terms of equivalence relations

$$\begin{aligned} R_{\mathcal{Y}^i \vee \mathcal{Y}^j} = & R_{\mathcal{Y}^i} \cup R_{\mathcal{Y}^i} \circ R_{\mathcal{Y}^j} \cup R_{\mathcal{Y}^i} \circ R_{\mathcal{Y}^j} \circ R_{\mathcal{Y}^i} \cup \dots \\ & \cup R_{\mathcal{Y}^j} \cup R_{\mathcal{Y}^j} \circ R_{\mathcal{Y}^i} \cup R_{\mathcal{Y}^j} \circ R_{\mathcal{Y}^i} \circ R_{\mathcal{Y}^j} \cup \dots \end{aligned} \quad (27)$$

where $R_x \circ R_y$ denotes the composition of the equivalence relations R_x and R_y and is the smallest equivalence relation containing both $R_{\mathcal{Y}^i}$ and $R_{\mathcal{Y}^j}$. $\mathcal{I}$ and $\mathcal{S}$ are defined in (6) and (7). Firstly, we prove that the $\mathcal{RF}$-product yields an equivalence relation

Theorem 1. *Let $R_{\mathcal{Y}^i}$ and $R_{\mathcal{Y}^j}$ be equivalence relations on a set U. Then $E = R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}$ is an equivalence relation on U.*
Proof. *E is an equivalence relation iff (1) $\forall E_i, E_j \in E, E_i \cap E_j = \emptyset$ and (2) $\cup E = U$.*

1. *Given that $R_{\mathcal{Y}^i}$ and $R_{\mathcal{Y}^j}$ are equivalence relations, $\mathcal{Y}^i$ and $\mathcal{Y}^j \in \Pi(U)$. $\forall u \in U, \exists R \in \mathcal{Y}^i$ and $\exists T \in \mathcal{Y}^j$ such that $u \in R$ and $u \in T$. Then $u \in R \cap T$. If $u \in R \cap T$ then $u \notin P \cap Q, \forall P \in \mathcal{Y}^i (P \neq R)$ and $\forall Q \in \mathcal{Y}^i (Q \neq T)$. The union of the intersections in the $\mathcal{RF}$-product ensure that u belongs to a single equivalence class $E_i \in E$ and hence $\forall E_i E_j \in E, E_i \cap E_j = \emptyset$.*
2. *Given that $\forall u \in U, \exists E_i \in E$ such that $u \in E_i$. Then $\cup E_i = U$.*

The operation $\mathcal{RF}$-product is commutative:

Theorem 2. *Let $R_{\mathcal{Y}^i}$ and $R_{\mathcal{Y}^j}$ be equivalence relations on a set U. Then $R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j} = R_{\mathcal{Y}^j} \otimes R_{\mathcal{Y}^i}$.*

Proof. *The property can be easly proven by first noting that intersection is a commutative operation. The matrix representing the $\mathcal{RF}$–product is built row-wise in $R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}$ (that is each row is the refinement of an equivalence class of $R_{\mathcal{Y}^j}$ by all equivalence classes of $R_{\mathcal{Y}^i}$), while it is built column-wise in $R_{\mathcal{Y}^j} \otimes R_{\mathcal{Y}^i}$ (that is each column is the refinement of an equivalence class of $R_{\mathcal{Y}^j}$ by all equivalence classes of $R_{\mathcal{Y}^i}$). In both cases the positions considered at the union step are the same, thus yielding the same result.*

Next we prove two theorems which bound the level of refinement of the partitions induced by the $\mathcal{RF}$-product.

Theorem 3. *Let $R_{\mathcal{Y}^i}$ and $R_{\mathcal{Y}^j}$ be equivalence relations on a set U. It holds that $R_{\mathcal{Y}^i} \cap R_{\mathcal{Y}^j} \subseteq R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}$.*

Proof. *From Eq. 17 it can be easily seen that each equivalence class of $R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}$ is the union of some equivalence classes of $R_{\mathcal{Y}^i} \cap R_{\mathcal{Y}^j}$. Then each equivalence class of $R_{\mathcal{Y}^i} \cap R_{\mathcal{Y}^j}$ is contained in an equivalence class of $R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}$. Hence $R_{\mathcal{Y}^i} \cap R_{\mathcal{Y}^j} \subseteq R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}$.*

Theorem 4. *Let $R_{\mathcal{Y}^i}$ and $R_{\mathcal{Y}^j}$ be equivalence relations on a set U. It holds that $(R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}) \otimes (R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}) = R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}$.*

Proof. *From Theorem 1 $R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}$ is an equivalence relation. Then $(R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}) \cap (R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}) = R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}$ and from Eq. 17 it derives that each equivalence relation of $R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}$ is equal to only one equivalence relation of $(R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}) \cap (R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j})$.*

Another interesting property of the $\mathcal{RF}$-product is that partition $E = R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}$ can be seen as the coarsest partition with respect to the sequence of operations

$$
\begin{aligned}
E &= R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j} \\
E' &= E \otimes R_{\mathcal{Y}^j} = (R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}) \otimes R_{\mathcal{Y}^j} \subseteq E \\
E'' &= E \otimes R_{\mathcal{Y}^i} = (R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}) \otimes R_{\mathcal{Y}^i} \subseteq E
\end{aligned}
$$

In other words, at each iteration, the $\mathcal{RF}$-product produces a finer partition with respect to the initial partition. It is worth noting that, at each iteration

$$E = E' \otimes E'' \tag{28}$$

Viewed from another perspective, the $\mathcal{RF}$-product can be seen as a rule generation mechanism. Suppose that it is possible to assign a label to each equivalence class of a partition. Then $R_{\mathcal{Y}^i} \otimes R_{\mathcal{Y}^j}$ represents a partition whose equivalence classes are consistent with the labels of the operands. Consider the following partitions on a set U

$$
\begin{aligned}
\mathcal{Y}^1 &= \{Y^1_{low}, Y^1_{medium}, Y^1_{high}\} \\
\mathcal{Y}^2 &= \{Y^2_{low}, Y^2_{medium}, Y^2_{high}\}
\end{aligned}
\tag{29}
$$

where $low = 1$ $medium = 2$ $high = 3$ and

$$
\begin{aligned}
\{u_4\} &= Y^1_{low} & \{u_3, u_4\} &= Y^2_{low} \\
\{u_2\} &= Y^1_{medium} & \{u_2\} &= Y^2_{medium} \\
\{u_1, u_3\} &= Y^1_{high} & \{u_1\} &= Y^2_{high}
\end{aligned}
$$

Applying $\mathcal{RF}$-product we get

$$\mathcal{Y}^{1,2} = \{Y^{1,2}_{low}, Y^{1,2}_{medium/low}, Y^{1,2}_{medium}, Y^{1,2}_{medium/high}, Y^{1,2}_{high}\} \tag{30}$$

where

$$
\begin{aligned}
&Y^{1,2}_{low} = \{(Y^1_{low} \cap Y^2_{low})\} \\
&Y^{1,2}_{medium/low} = \{(Y^1_{medium} \cap Y^2_{low}) \cup (Y^1_{low} \cap Y^2_{medium})\} \\
&Y^{1,2}_{medium} = \{(Y^1_{high} \cap Y^2_{low}) \cup (Y^1_{medium} \cap Y^2_{medium}) \cup (Y^1_{low} \cap Y^2_{high})\} \\
&Y^{1,2}_{medium/high} = \{(Y^1_{high} \cap Y^2_{medium}) \cup (Y^1_{medium} \cap Y^2_{high})\} \\
&Y^{1,2}_{high} = \{(Y^1_{high} \cap Y^2_{high})\}
\end{aligned}
\tag{31}
$$

Analyzing the new partition we note how the equivalence classes are consistent with the composition of the original ones, i.e.:

a) $u \in U$ belongs to "low" class in $Y^{1,2}$ if it belongs to "low" class in Y^1 and "low" class in Y^2;
b) $u \in U$ belongs to "medium/low" class in $Y^{1,2}$ if it belongs to "low" class in Y^1 and "medium" class in Y^2 or to "medium" class in Y^1 and "low" class in Y^2;
c) $u \in U$ belongs to "medium" class in $\mathcal{Y}^{1,2}$ if it belongs to "medium" class in Y^1 and Y^2 or to "high" class in Y^1 and to "low" class in Y^2 or to "high" class in Y^2 and to "low" class in Y^1;
d) $u \in U$ belongs to "medium/high" class in $Y^{1,2}$ if it belongs to "high" class in Y^1 and "medium" class in Y^2 or to "medium" class in Y^1 and "high" class in Y^2;
e) $u \in U$ belongs to "high" class in $Y^{1,2}$ if it belongs to "high" class in Y^1 and Y^2.

5 Feature Discovery

The basic ideas about how to construct the feature vectors upon the definitions introduced in the previous section are outlined for the case of image analysis.

Let us consider an image I defined over a set $U = [0, ..., H-1] \times [0, ...W-1]$ of picture elements, i.e. $I : u \in U \rightarrow [0, 1]$. Let us also consider a grid, superimposed on the image, whose cells Y_i are of dimension $w \times w$, such that all Y_i constitute a partition over I, i.e. eqs (24) and (25) hold and each Y_i^1, for $i = 1 \dots p$, has dimension $w \times w$ and $p = H/w + W/w$. The size w of each equivalence class will be referred to as *scale*.

Each cell of the grid can be seen as an equivalence class induced by an equivalence relation $\mathcal{R}$ that assigns each pixel of the image to a single cell. Given a pixel u, whose coordinates are u_x and u_y, and a cell Y_i of the grid, whose coordinates of its upper left point are $x(Y_i)$ and $y(Y_i)$, u belongs to Y_i if $x(Y_i) \leq u_x \leq x(Y_i)+w-1$ and $y(Y_i) \leq u_y \leq y(Y_i)+w-1$. In other words, we are defining a partition U/R of the image induced by the relation $\mathcal{R}$, in which each cell represents an equivalence class $[u]_{\mathcal{R}}$. Also suppose that equivalence classes can be ordered in some way, for instance, from left to right.

Moreover, given a subset X of the image, not necessarily included or equal to any $[u]_{\mathcal{R}}$, we define the membership degree $\mu_X(u)$ of a pixel u to X as the normalized gray level value of the pixel.

If we consider different scales, the partitioning scheme yields many partitions of the same image and hence various approximations $\overline{R}(X)$ and $\underline{R}(X)$ of the subset X. For instance, other partitions can be obtained by a rigid translation of $\mathcal{Y}^1$ in the directions of 0^o, 45^o and 90^o of $w-1$ pixels, so that for each partition a pixel belongs to a shifted version of the same equivalence class Y_j^i.

If we consider four equivalence classes, Y_j^1 Y_j^2 Y_j^3 Y_j^4 as belonging to these four different partitions, then there exists a pixel u with coordinates u_x, u_y such that u belongs to the intersection of Y_j^1 Y_j^2 Y_j^3 Y_j^4. Hence each pixel can be seen as belonging to the equivalence class

$$Y_j^{1,2,3,4} = Y_j^1 \cap Y_j^2 \cap Y_j^3 \cap Y_j^4 \tag{32}$$

of the partition obtained by $\mathcal{RF}$-producting the four rough fuzzy sets to which Y_j^i, with $i = 1, \ldots, 4$, belongs, i.e.

$$RF_X^{1,2,3,4} = RF_X^1 \otimes RF_X^2 \otimes RF_X^3 \otimes RF_X^4 \tag{33}$$

The $\mathcal{RF}$-product behaves as a filtering process according to which the image is filtered by a minimum operator over a window $w \times w$ producing $\mathcal{I}$ and by a maximum operator producing $\mathcal{S}$. Iterative application of this procedure consists in applying the same operator to both results $\mathcal{I}$ and $\mathcal{S}$ obtained at the previous iteration.

As instance, X defines the contour or uniform regions in the image. On the contrary, regions appear rather like fuzzy sets of grey levels and their comparison or combination generates more or less uniform partitions of the image. Rough fuzzy sets, as defined in (5), seem to capture these aspects together, trying to extract different kinds of knowledge in data.

This procedure can be efficiently applied to image coding/decoding, getting rise to the method *rough fuzzy vector quantization* (RFVQ)[12]. The image is firstly partitioned in non–overlapping k blocks X_h of dimension $m \times m$, such that $m \geq w$, that is $X = \{X_1, \ldots, X_k\}$ and $k = H/m + K/m$.

Considering each image block X_h, a pixel in the block can be characterized by two values that are the membership degrees to the lower and upper approximation of the set X_h. Hence, the feature extraction process provides two approximations $\underline{R}(X_h)$ and $\overline{R}(X_h)$ characterized by $\mathcal{I}$ and $\mathcal{S}$ as defined in (6) and (7) where

$$\begin{aligned} \underline{\nu}_i &= \mu_{\underline{R}(X_h)}(Y_i) = \inf\{\mu_{X_h}(u) | Y_i = [u]_R\} \\ \overline{\nu}_i &= \mu_{\overline{R}(X_h)}(Y_i) = \sup\{\mu_{X_h}(u) | Y_i = [u]_R\} \end{aligned} \tag{34}$$

and $[u]_R$ is the granule that defines the resolution at which we are observing the block X_h. For a generic pixel $u = (u_x, u_y)$ we can compute the coordinates of the upper left pixel of the four equivalence classes containing u, as shown in Fig. 1:

$$
\begin{aligned}
u_x &= x_1 + w - 1 \Rightarrow x_1 = u_x - w + 1 \\
u_y &= y_1 + w - 1 \Rightarrow y_1 = u_y - w + 1 \\
u_x &= x_2 \Rightarrow x_2 = u_x \\
u_y &= y_2 + w - 1 \Rightarrow y_2 = u_y - w + 1 \\
u_x &= x_3 + w - 1 \Rightarrow x_3 = u_x - w + 1 \\
u_y &= y_3 \Rightarrow y_3 = u_y \\
u_x &= x_4 \Rightarrow x_4 = u_x \\
u_y &= y_4 \Rightarrow y_4 = u_y
\end{aligned}
$$

where the four equivalence classes for pixel u are

$$
\begin{aligned}
Y_j^1 &= (x_1, y_1, \underline{\nu}_j^1, \overline{\nu}_j^1) \\
Y_j^2 &= (x_2, y_2, \underline{\nu}_j^2, \overline{\nu}_j^2) \\
Y_j^3 &= (x_3, y_3, \underline{\nu}_j^3, \overline{\nu}_j^3) \\
Y_j^4 &= (x_4, y_4, \underline{\nu}_j^4, \overline{\nu}_j^4)
\end{aligned}
$$

For instance, if we choose a granule of dimension $w = 2$ for a generic $j - th$ granule of the $i - th$ partition, equations in (34) become:

$$
\begin{aligned}
\underline{\nu}_j^i &= \inf\{(u_x + a, u_y + b) | a, b = 0, 1\} \\
\overline{\nu}_j^i &= \sup\{(u_x + a, u_y + b) | a, b = 0, 1\}
\end{aligned}
$$

The compression method performed on each block X_h is composed of three phases: *codebook design*, *coding* and *decoding*. A vector is constructed by retaining the values $\underline{\nu}_j^i$ and $\overline{\nu}_j^i$ at positions u and $u + (w - 1, w - 1)$ in a generic block X_h, or equivalently $\underline{\nu}_j^1, \overline{\nu}_j^1, \underline{\nu}_j^3, \overline{\nu}_j^3$. The vector has hence dimension m^2 consisting of $m^2/2$ inf values and $m^2/2$ sup values. The vectors so constructed and extracted from a training image set are then fed to a quantizer in order to construct the codebook. The aim of vector quantization is the representation of a set of vectors $u \in X \subseteq R^{m^2}$ by a set of C prototypes (or codevectors) $V = \{v_1, v_2, ..., v_C\} \subseteq R^{m^2}$. Thus, vector quantization can also be seen as a mapping from an m^2-dimensional Euclidean space into the finite set $V \subseteq R^{m^2}$, also referred to as the codebook. Codebook design can be performed by clustering algorithms, but it is worth noting that the proposed method relies on the representation capabilities of the vector to be quantized and not on the quantization algorithm, to determine optimal codevectors, i.e. Fuzzy C-Means, Generalized Fuzzy C-Means or any analogous clustering algorithm can be adopted.

The process of coding a new image proceeds as follows. For each block X_h the features extracted are arranged in a vector, following the same scheme used for designing the codebook, and compared with the codewords in the codebook to find the best match, i.e. the closest codeword to the block.

In particular, for each block, inf and sup values are extracted from a window of size 2×2 shifted by one pixel into the block. All the extracted values are arranged in a one-dimensional array, i.e. for block dimension $m \times m$ and a window dimension 2×2 the array is represented by m^2 elements consisting of $m^2/2$ inf values and $m^2/2$ sup values. Doing so, the identificative number (out of C) of the winning codeword, i.e. the best match to the coded block, is saved in place of the generic block X_h.

Given a coded image, the decoding step firstly consists in the substitution of the identificative codeword number with the codeword itself, as reported in the codebook. The codeword consists of $m^2/2$ inf and $m^2/2$ sup values, instead of the original m^2 values of the block. To reconstruct the original block, we apply the theory as follows. As stated above, each pixel can be seen as belonging to the block of the partition obtained by $\mathcal{RF}$-producting the four equivalence classes (32) and (33). Specifically, the blocks contained into the codeword are not the original ones, but those chosen to represent the block, i.e.

$$Y_j^{1,2,3,4} = Y_j^1 \cap Y_j^2 \cap Y_j^3 \cap Y_j^4 \tag{35}$$

where Y_j^i is a set of the partition of the rough-fuzzy set intersecting the generic block of the image X_h. The result of the $\mathcal{RF}$-product operation, with respect to a single block, is represented by another rough fuzzy set, characterized by lower and upper approximations. These values are used to fill the missing values into the decoded block.

In detail, being q_r the codeword corresponding to a generic block, the decoded block $X_{decoded}$ is constructed by filling the missing values, i.e. the original $\underline{\nu}_j^2, \overline{\nu}_j^2, \underline{\nu}_j^4, \overline{\nu}_j^4$ as combination of them, like average, median, etc., yielding $\tilde{\underline{\nu}}_j^2, \tilde{\overline{\nu}}_j^2, \tilde{\underline{\nu}}_j^4, \tilde{\overline{\nu}}_j^4$.

The reconstructed block $X_{decoded}$, again a rough fuzzy set, is obtained by $\mathcal{RF}$-producting the four equivalnce classes Y_j^1 Y_j^2 Y_j^3 Y_j^4, yielding the following

$$\begin{aligned}
Y_j^{1,2,3,4} &= Y_j^1 \cap Y_j^2 \cap Y_j^3 \cap Y_j^4 \\
\tilde{\mathcal{I}}^{1,2,3,4}(u) &= \sum_j \tilde{\underline{\nu}}_j^{1,2,3,4} \times \mu_{Y_j}^{1,2,3,4}(u) \\
\tilde{\mathcal{S}}^{1,2,3,4}(u) &= \sum_j \tilde{\overline{\nu}}_j^{1,2,3,4} \times \mu_{Y_j}^{1,2,3,4}(u)
\end{aligned}$$

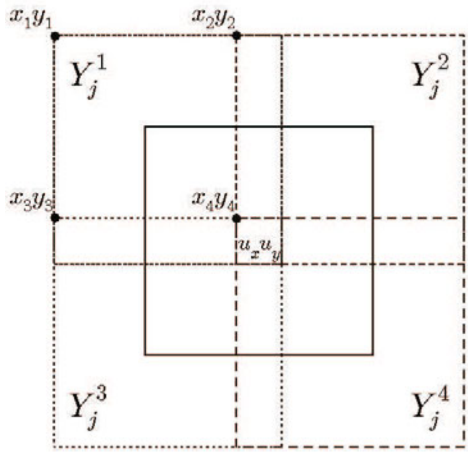

Fig. 1 Equivalence classes coordinates.

where

$$\tilde{\underline{\nu}}_j^{1,2,3,4} = \sup\{\underline{\nu}_j^1, \tilde{\underline{\nu}}^2{}_j, \underline{\nu}_j^3, \tilde{\underline{\nu}}_j^4\} \qquad \tilde{\overline{\nu}}_j^{1,2,3,4} = \inf\{\overline{\nu}_j^1, \tilde{\overline{\nu}}_j^2, \overline{\nu}_j^3, \tilde{\overline{\nu}}_j^4\} \tag{36}$$

Lastly, under the assumption of local smoothness an estimate of the original grey values at the generic position u can be computed composing $\tilde{\underline{\nu}}_j^{1,2,3,4}$ and $\tilde{\overline{\nu}}_j^{1,2,3,4}$, as instance averaging or simply using only one of them.

An example of compression of image *Baboon* is depicted in Figure 2(a) using RFVQ. The codebook has been built by using ISODATA quatizer, a training image set of size 256×256 with 8 bits/pixel: Bird, Bridge, Building, Camera, City, Hat, House, Lena, Mona, Salesman, using the average as combination. Different compression rates are obtained as explained in Table 2.

Table 2 RFVQ parameters at different compression rates.

Compression rate	Number of clusters (C)	Block dimension (m)
0.03	16	4
0.06	256	4
0.14	32	2
0.25	256	2
0.44	16384	2

Analyzing the results shown in Fig.s 2 (b)–(f), we can observe that the proposed method performs well for higher compression rates while it looses efficiency for lower compression rates, reasonably due to the quantization algorithm. Indeed, in order to obtain a compression rate of 0.44 a large number of clusters has to be computed (precisely 16,384), but in this situation the large number of codewords does not ensure that the optimal choice will be performed when selecting the most approximating codeword.

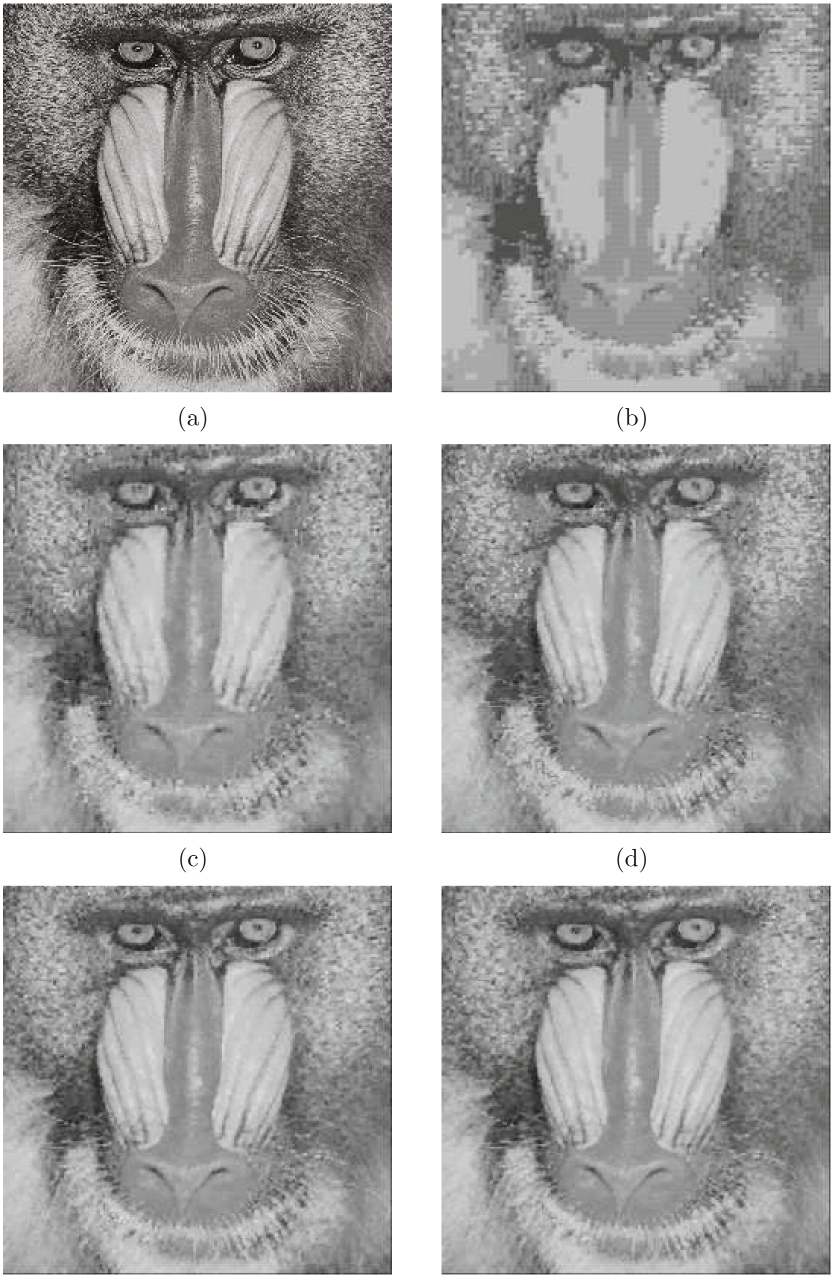

Fig. 2 Image "Baboon" (a) at different compression rates, b) 0.03 (PSNR: 19.10), c) 0.06 (PSNR: 19.89), d) 0.14 (PSNR: 20.61), e) 0.25 (PSNR: 21.08), f) 0.44 (PSNR: 21.36).

6 Concluding Remarks

A model for the hybridization of rough and fuzzy sets has been presented. It is endowed with a new operator, called $\mathcal{RF}$-product, which allows to combine different partitions yielding a refined partition in a hierarchical manner. The model and the $\mathcal{RF}$-operator have been proved to possess peculiar properties effective for feature discovery, useful in applications like image analysis. Here we reported an image compression scheme that exploits the peculiarities of $\mathcal{RF}$-product. Results are quite remarkable, considering that RFVQ does not suffer of the blocking effect, while loosing only a small amount of details. Ongoing work is devoted to exploit the proposed model in other data mining applications and also image processing tasks, like color image segmentation.

References

[1] Sen, D., Pal, S.K.: Generalized Rough Sets, Entropy, and Image Ambiguity Measures. IEEE Trans. on Syst. Man and Cybern. 39, 117–128 (2009)
[2] Caianiello, E.R.: A calculus of hierarchical systems. In: Proc. Internat. Conf. on Pattern Recognition, pp. 1–5 (1973)
[3] Caianiello, E.R., Petrosino, A.: Neural networks, fuzziness and image processing. In: Cantoni, V. (ed.) Machine and Human Perception: Analogies and Divergencies, Plenum Press, NewYork (1994)
[4] Caianiello, E.R., Ventre, A.: A model for C-calculus. International Journal of General Systems 11, 153–161 (1985)
[5] Yao, Y.Y.: Information granulation and rough set approximation. International Journal of Intelligent Systems 16, 87–104 (2001)
[6] Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. Int. J. Gen. Sys. 17, 191–209 (1990)
[7] Li, Y., Qiao, B.: Hierarchical Reduction Algorithm of Rough Sets. In: Sixth International Conference on Intelligent Systems Design and Applications, vol. 1, pp. 497–502 (2006)
[8] Qian, Y., Liang, J., Yao, Y., Dang, C.: MGRS: A multi-granulation rough set. Information Sciences 180, 949–970 (2010)
[9] Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 11, 341–356 (1982)
[10] Pawlak, Z.: Granularity of knowledge, indiscernibility and rough sets. In: Proceedings of IEEE International Conference on Fuzzy Systems, pp. 106–110 (1998)
[11] Petrosino, A., Salvi, G.: Rough fuzzy set based scale space transforms and their use in image analysis. International Journal of Approximate Reasoning 41, 212–228 (2006)
[12] Petrosino, A., Ferone, A.: Rough fuzzy set-based image compression. Fuzzy Sets and Systems 160, 1485–1506 (2009)
[13] Zadeh, L.: Fuzzy sets. Information and Control, pp. 338–353 (1964)