A Preliminary Approach to Referring to Groups of Objects in Images

Nicolás Marín, Gustavo Rivas-Gervilla*, and Daniel Sánchez
Dept. Computer Science and A.I. and CITIC-UGR
University of Granada
Granada, Spain
email: {nicm,griger,daniel}@decsai.ugr.es

***Abstract*—In order to talk about an image it is of special interest to be able to refer to the groups of objects that appear in it. To do that, we have to form noun phrases in which properties are combined to discriminate the target group from the rest. These phrases are part of what is known as plural referring expressions in the NLG field. The construction of this type of expressions is a complex problem, given the graduality of both the visual properties that can be used to make them up and the concept of group itself, among other reasons. In this work we propose the use of distance as the inducing element of the notion of group of objects in the image and we also show how results from the theory of fuzzy sets and the related theory of representation by levels (RL) can be used to assess the referential success of referring expressions taking into account the aforementioned graduality of the involved concepts.**

***Index Terms*—Object Group Identification, Referring Expression Generation**

I. Introduction

The identification of groups of objects in images attempts to determine the different natural groups humans are able to distinguish when inspecting an image. This is a key task in different applications, particularly in data-to-text systems, including image description and object location, among others.

There are many factors that can influence the identification of object groups. One obvious factor is distance among objects. A reasonable requirement for a set of objects to be an object group is that they are closer to each other than to objects in other groups. In this sense, an obvious approach to identification of object groups is to perform a clustering process, by computing similarity on the basis of distance. As an example, consider the image in Fig. 1. In this scene we can identify two clear groups according to the distance between the objects. These groups are the set of objects $A = \{t_1, s_1, c_4\}$ and $B = \{c_1, c_2, c_3, t_3\}$. The rest of objects are more dispersed in the image, so attending only to the distance we need to relax (sometimes considerably) our criterion of "vicinity" in order to identify other groups in this context. For example, looking at the image, it also seems clear that the set A of three objects is perceived as a group slightly more easily than the set B of four objects.

This example shows an important idea of how groups are formed: distance is the main criterion when determining groups and its influence is a matter of degree, some groups being easier to identify than others. We should call *salience* to the degree to which it is easy to identify a group [1]. Different approaches to define salience are available in the literature, see for example [2]–[4].

In data-to-text applications, once visually salient groups of objects are identified, it is important to generate linguistic expressions able to distinguish them from other groups [5]–[7]. These expressions are called *referring expressions* [1], [8]. In the particular case of groups of objects, we are concerned with the problem of *referring to sets* or *plural referring expression generation*, as particular cases of the more general referring expression generation (REG) problem. We shall use REG for simplicity along the paper when referring to the plural version of this problem. As related work outside the NLG community, the problem of identifying object groups has also been considered in the setting of flexible querying in fuzzy databases in order to characterize the answer of queries [9]–[11].

As an example, consider again our previous image in Fig. 1. Regarding some of the groups we have identified before, we can refer to group B as "the group of gray objects". On its turn, we can refer to group A as "the group of white objects that contains a circle, a triangle and a square". These two noun phrases are examples of referring expressions. Note that it is possible in general to find different referring expressions for the same group. For instance, we can also refer to group B as "the group that contains a gray triangle"; similarly, we can also refer to group A as "the group of objects that contains a white square".

The objective of this paper is to offer a preliminary approximation to the problem of referring to groups of objects in images in a fuzzy setting, which opens a research line about REG for groups of objects in images using fuzzy properties. Fuzziness come into play in several aspects of this problem: the identification of groups of objects (related to the notion of salience we have mentioned before), the properties employed in referring expressions, and the assessment of

*Corresponding author.
This work has been partially supported by the Spanish Ministry of Science, Innovation and Universities and the European Regional Development Fund - ERDF (Fondo Europeo de Desarrollo Regional - FEDER) under project PGC2018-096156-B-I00 *Recuperación y Descripción de Imágenes mediante Lenguaje Natural usando Técnicas de Aprendizaje Profundo y Computación Flexible*. This work has been also partially supported by the Spanish Ministry of Education, Culture and Sports grant FPU16/05199.

978-1-7281-6932-3/20/$31.00 ©2020 IEEE

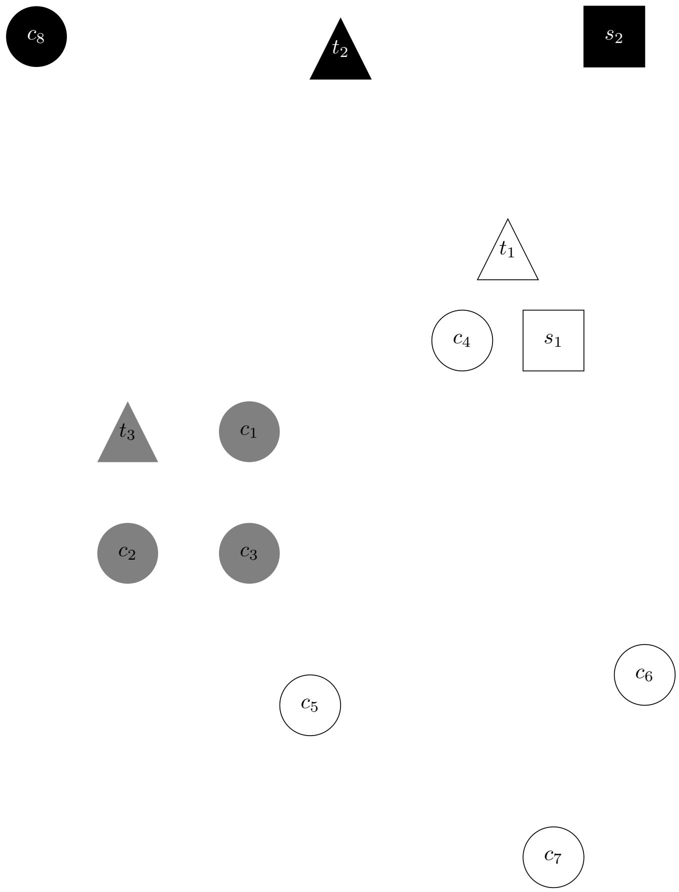

Fig. 1. Scene example.

referential success in the generation of referring expressions. More specifically:

- It is well known that similarity in terms of distance is a matter of degree.
- Many visual features that can be considered in referring expressions, like those related to shape, texture, and specially color, are also considered as paradigmatic examples of fuzzy concepts [12].
- With respect to the REG problem, the suitability of a referring expression for distinguishing a group, called *referential success*, becomes a matter of degree as well [12]–[17]. For instance, in the image in Fig. 1, the expression "the group of white objects" could be identified as either group A or group $C = \{c_5, c_6, c_7\}$. Though objects in group C are not so close to each other as those in group A, they are far enough from other objects so that they can be identified as a group on the basis of distance only, but with more difficulties. Even considering that group C is less salient that group A, its presence in the image can affect the referential success of the expression "the group of white objects" for referring to group A, as we may hesitate between A and C as the target group of the referring expression. Note also that the salience of a group $D = \{t_1, s_1, c_4, c_5, c_6, c_7\}$ is naturally much lower than those of groups A and C due to the distance between the objects, to the point that it is much less likely in general that a human would consider group D as the group we are referring to.[1]

The paper is organized as follows: we discuss on distance-based object group identification in Section II. Our proposal of properties and referential success measures is described in Section III, together with a discussion about referring expression generation. Finally, section IV contains our conclusions and future research lines.

II. Distance-based object group identification

We assume in this section that individual objects in the image have been identified by means of some suitable image analysis technique. Together with the objects, we assume that a measure of distance between every pair of objects has been calculated, either in the form of distance between object centres, smaller distance between pixels of every object, or using distance between the corresponding bounding boxes associated to the objects. Any distance can be employed, provided it is suitable for the purpose of discriminating groups of objects. In our examples, we use synthetic images with simple scenes in which both objects and the distance between every pair of objects are available.

As we suggested in the introduction, we shall use clustering with similarity based on distance for basic determination of object groups. Hence, we start from a set $O = \{o_1, \ldots, o_n\}$ of objects, $n \geq 2$, and a distance $d(\cdot, \cdot)$ between objects in O. We shall employ coverings and a multi-level perspective, since groups can be appreciated at different scales and, when we focus on the context given by the objects of a single group, it is usual to observe subgroups of objects, a procedure that can be repeated recursively until we arrive to individual objects. In the next sections we shall explain our clustering procedure and we shall propose a measure of salience with the objective or determining how "easy" is to perceive each of the obtained groups.

A. Finding groups using coverings by levels

As it is usual in many applications, we start by computing a resemblance (reflexive and symmetric) fuzzy binary relation R^{d_m} on O on the basis of distance, as follows:

$$R^{d_m}(o_i, o_j) = \max\left\{1 - \frac{d(o_i, o_j)}{d_m}, 0\right\} \quad (1)$$

where $d_m > 0$ is a parameter that defines the distance beyond which the resemblance between objects is 0. This parameter determines in our clustering procedure the maximum diameter of a sphere containing all the objects in a group. One possibility is to define d_m as the maximum diameter of the area of the image that we can see in a single sight, but any

[1] In this point, note the difference between "the group of white objects" and "the set of white objects". The latter can easily be applied to D since the term "set" does not imply an spatial connotation; on the contrary, the term "group" implies that objects are close to each other and separate from objects in other groups, as we have explained before.

other value can be used depending on the requirements of the specific application at hand. In the following we shall assume that d_m is determined and known, so that we shall denote R^{d_m} simply as R.

On the basis of such resemblance relation, a clustering procedure is proposed in [18] based on the representation by levels (RLs) [19], which is an alternative to fuzzy sets as a tool for representing and operating with sets having ill-known boundaries[2]. The main features of RLs are: i) every crisp mathematical object (set, number, etc.) is taken to the fuzzy case in an unique and easy way, as an assignment of objects to levels in (0,1], and ii) operations between RLs are performed in every level independently, results keeping all the properties of the crisp case. Particularly, RLs of sets form a Boolean algebra with respect to union, intersection, and complementation. Similarly, RLs of numbers (natural, integer, real, etc.) keep the same properties and algebraic structure of their crisp counterparts.

In the case of clustering, the procedure proposed in [18] provides an assignment of crisp coverings to levels in the set $\Lambda_R = \{\alpha \in (0,1]$ such that $\exists o_i, o_j \in O$ with $R(o_i, o_j) = \alpha\}$. It is immediate that $\Lambda_R = \{\alpha_1, \ldots, \alpha_k\}$ with $k \geq 1$ and $1 = \alpha_1 > \cdots > \alpha_{k+1} = 0$. For each $\alpha_k \in \Lambda_R$, a crisp covering is obtained as follows:

1) Obtain R_{α_i} (the α_i-cut of R).
2) The set of maximal cliques in R_{α_i} defines a crisp covering of O in level α_i with k_i (non-disjoint) clusters $\mathcal{C}_i = \{C_{i1}, \ldots, C_{ik_i}\} \subset \{0,1\}^O$.

Then, a ***RL-clustering*** based on coverings is obtained as a pair (Λ_R, ρ_R), where ρ_R is a function assigning crisp coverings to levels, i.e., for every level $\alpha_i \in \Lambda_R$ it is

$$\rho_R(\alpha_i) = \mathcal{C}_i \tag{2}$$

By the properties of maximal cliques of crisp binary relations, it is also easy to prove that for every $1 < i < k$ and for every cluster C in level α_i, there are at least one cluster C' in level α_{i-1} and at least one cluster C'' in level α_{i+1} such that $C' \subseteq C \subseteq C''$. That is, the collection of coverings corresponding to the different levels has a structure with k levels, indexed from 1 (where all sets are singletons, since this level consider that two objects are similar when the distance between them is 0 only) to k. We shall denote by $\mathcal{G}$ the collection of all distinguishable groups of objects in the image:

$$\mathcal{G} = \bigcup_{i=1}^{k} \mathcal{C}_i = \{g_1, \ldots, g_l\} \tag{3}$$

That is, $\mathcal{G}$ is the set containing all the clusters obtained at each level of the structure[3].

[2]This clustering procedure has been applied for instance to solve the context-dependent semantics of fuzzy sets modeling size [20].

[3]Note that this deterministic procedure is not a conventional Agglomerative Hierarchical Clustering, since i) in each level, it produces in general crisp clusterings comprised of non-disjoint clusters, not partitions, and ii) it builds a final collection of groups of objects $\mathcal{G}$ that it is not a clustering of the objects in the image, since some groups are contained in others when the number of objects is greater than 1.

As an example Tables II and III show the RL-clustering obtained for the image in Figure 1 using $d_m = 6$ and $d_m = 4$, respectively, where levels have been rounded to the nearest two decimal values. Euclidean distances between objects are shown in Table I.

B. Distance-based salience

For every $g_i \in \mathcal{G}$, we can define

$$m(g_i) = \sum_{\alpha_j \in \Lambda_R \ | \ g_i \in \rho_R(\alpha_j)} (\alpha_j - \alpha_{j+1}) \tag{4}$$

In the RL theory, the measure m in Eq. (4) is the probability that a level α taken at random in $(0,1]$ satisfies $g_i \in \mathcal{C}_i = \rho_R(\alpha_i)$, with $\alpha_i = \min\{\alpha_j \in \Lambda_R$ such that $\alpha_j \geq \alpha\}$, hence $m(g_i) \in [0,1]$. In our context, this value can be seen as the degree to which it is easy to distinguish g_i among the groups in $\mathcal{G}$.

Using this measure, we can define the *distance-based salience* for every group $g_i \in \mathcal{G}$ as a value in $[0,1]$ obtained by normalizing the measure m in $\mathcal{G}$, that is

$$sal(g_i) = \frac{m(g_i)}{M} \tag{5}$$

where

$$M = \max_{g_i \in \mathcal{G}} \ m(g_i). \tag{6}$$

For the sake of simplicity, we shall use the term *salience* to refer to the *distance-based salience* in the rest of the paper.

As an example Tables IV and V show values of m and salience for some of the groups in Tables II and III, respectively, also rounded to the nearest real number with a precision of two decimals.

III. Referring expressions for groups of objects

From the application of techniques like the one described in the previous section, we obtain a set of object groups $\mathcal{G}$ which are partially ordered according to the classical inclusion relationship between sets; that is, $\mathcal{G}$ is a poset. As we will see in this section, this poset can serve as a basis for obtaining referring expressions that allow identifying groups of objects in the image.

A referring expression is a noun phrase whose aim is to univocally identify an element within a collection. In the case of this paper, the aim of referring expressions is to univocally identify groups of objects within $\mathcal{G}$. Usually, a referring expression takes the form of a conjunction of properties.

A. Properties

In this paper we will consider two types of properties when describing groups and constructing plural referring expressions:

- *Collective properties* are properties of a set of objects as a whole. These properties can have different nature. In this work we are going to consider the *c*ardinality of the set (e.g., "to have three objects"), and those properties that all elements of the set share and that can be represented by the template (e.g. "all its objects are gray"). Within

TABLE I
DISTANCES BETWEEN OBJECTS IN FIGURE 1.

	c_2	t_3	c_3	c_1	c_4	s_1	t_1	c_5	c_6	c_7	s_2	t_2	c_8
c_2	0												
t_3	2	0											
c_3	2	3	0										
c_1	3	2	2	0									
c_4	7	6	5	4	0								
s_1	8	7	6	5	2	0							
t_1	8	7	7	5	2	2	0						
c_5	4	5	3	5	6	7	8	0					
c_6	9	9	7	8	6	6	7	6	0				
c_7	9	10	7	9	9	8	10	5	3	0			
s_2	12	10	10	9	6	5	4	12	11	14	0		
t_2	9	7	8	6	5	6	4	11	11	14	5	0	
c_8	9	7	9	7	9	10	9	12	14	16	10	5	0

TABLE II
RL-CLUSTERING OBTAINED FOR THE IMAGE IN FIGURE 1 WITH $d_m = 6$.

Distance	Level	Covering
0	1	$\{c_1\},\{c_2\},\{c_3\},\{c_4\},\{c_5\},\{c_6\},\{c_7\},\{c_8\},\{t_1\},\{t_2\},\{t_3\},\{s_1\},\{s_2\}$
2	0.67	$\{c_2,t_3\},\{c_2,c_3\},\{t_3,c_1\},\{c_3,c_1\},\{c_4,s_1,t_1\},\{c_5\},\{c_6\},\{c_7\},\{s_2\},\{t_2\},\{c_8\}$
3	0.5	$\{c_2,t_3,c_3,c_1\},\{c_3,c_5\},\{c_4,s_1,t_1\},\{c_6,c_7\},\{s_2\},\{t_2\},\{c_8\}$
4	0.33	$\{c_2,t_3,c_3,c_1\},\{c_2,c_3,c_5\},\{c_1,c_4\},\{c_4,s_1,t_1\},\{t_1,s_2\},\{t_1,t_2\},\{c_6,c_7\},\{c_8\}$
5	0.17	$\{c_2,t_3,c_3,c_1,c_5\},\{c_3,c_1,c_4\},\{c_1,c_4,s_1,t_1\},\{c_4,t_1,t_2\},\{s_1,t_1,s_2\},\{t_1,s_2,t_2\},\{c_5,c_7\},\{c_6,c_7\},\{t_2,c_8\}$

TABLE III
RL-CLUSTERING OBTAINED FOR THE IMAGE IN FIGURE 1 WITH $d_m = 4$.

Distance	Level	Covering
0	1	$\{c_1\},\{c_2\},\{c_3\},\{c_4\},\{c_5\},\{c_6\},\{c_7\},\{c_8\},\{t_1\},\{t_2\},\{t_3\},\{s_1\},\{s_2\}$
2	0.5	$\{c_2,t_3\}$, $\{c_2,c_3\}$, $\{t_3,c_1\}$, $\{c_3,c_1\}$, $\{c_4,s_1,t_1\}$, $\{c_5\}$, $\{c_6\}$, $\{c_7\}$, $\{s_2\}$, $\{t_2\}$, $\{c_8\}$
3	0.25	$\{c_2,t_3,c_3,c_1\}$, $\{c_3,c_5\}$, $\{c_4,s_1,t_1\}$, $\{c_6,c_7\}$,$\{s_2\}$,$\{t_2\}$, $\{c_8\}$

this kind of properties there is one of special interest: "to be a group".

- Additionally, we will consider properties derived from inclusion relationships as, for example, "to contain *a set* of four triangles". When the set of objects satisfy the property "to be a group" as well, the conjunction of both properties can be linguistically expressed as "to contain *a group* of four triangles".

Since the above mentioned properties can be fuzzy in general, in order to generate referring expressions using these properties, a first step is to determine how to assess their accomplishment degree.

Let us consider again an image with a set of objects $O = \{o_1, \ldots, o_n\}$ and a set of groups $\mathcal{G} = \{g_1, \ldots, g_l\}$. Let us consider a set $P = \{p_1, \ldots, p_m\}$ of properties that may be satisfied by objects in O.

Properties in P can be fuzzy properties, with the fulfilment degree of property $p_i \in P$ by an object $o \in O$ denoted by $p_i(o) \in [0,1]$. Let us formalize the following sets of properties that may be satisfied by any set of objects:

- We shall consider *cardinality* as a crisp property. Hence, the properties related to cardinality will be denoted as
$$Card = \{card_1, \ldots, card_n\} \tag{7}$$
where $card_i$ means "to have i objects", i.e., for every $s \in \{0,1\}^O$ it is
$$card_i(s) = \begin{cases} 1 & \text{when } |s| = i \\ 0 & \text{otherwise} \end{cases} \tag{8}$$
- The property gr, meaning "to be a group", is also a crisp property defined as follows: for every $s \in \{0,1\}^O$ it is
$$gr(s) = \begin{cases} 1 & \text{when } s \in \mathcal{G} \text{ and } |s| > 1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$
- In addition, we shall consider the set of collective properties
$$SP^P = \{sp_1^P, \ldots, sp_m^P\} \tag{10}$$
where sp_i^P means "all the objects of the set satisfy p_i", with $p_i \in P$. For every $s \in \{0,1\}^O$ we define
$$sp_i^P(s) = \min_{o \in s} p_i(o) \tag{11}$$
- We shall also consider a set of inclusion properties
$$P^{\subset} = \{p^{s,\Gamma} \mid s \in \{0,1\}^O \wedge \Gamma \in \{0,1\}^{SP^P \cup Card \cup \{gr\}}\} \tag{12}$$
where $p^{s,\Gamma}$ means "to include a set s that satisfies all properties in Γ". We define
$$p^{s,\Gamma}(s') = \begin{cases} 0 & \text{when } s \not\subset s' \\ \min_{\gamma \in \Gamma} \gamma(s) & \text{otherwise} \end{cases} \tag{13}$$

As a final remark, let us stress again the difference between inclusion properties defined using sets vs. those defined using

TABLE IV
VALUES OF m AND SALIENCE FOR SOME OF THE GROUPS IN TABLE II.

Group	m	Salience
$\{c_3\}$	0.33	0.4
$\{c_1, c_4\}$	0.17	0.2
$\{s_1\}$	0.33	0.4
$\{t_2\}$	0.67	0.8
$\{c_6\}$	0.5	0.6
$\{c_2, c_3\}$	0.17	0.2
$\{c_6, c_7\}$	0.5	0.6
$\{c_4, s_1, t_1\}$	0.5	0.6
$\{t_3, c_1\}$	0.17	0.2
$\{c_1, c_4, s_1, t_1\}$	0.17	0.2
$\{c_2, c_3, c_5\}$	0.17	0.2
$\{t_1, t_2\}$	0.17	0.2
$\{c_4\}$	0.33	0.4
$\{t_3\}$	0.33	0.4
$\{c_5\}$	0.5	0.6
$\{s_1, t_1, s_2\}$	0.17	0.2
$\{s_2\}$	0.67	0.8
$\{c_2, t_3\}$	0.17	0.2
$\{c_3, c_5\}$	0.17	0.2
$\{c_5, c_7\}$	0.17	0.2
$\{c_2, t_3, c_3, c_1\}$	0.33	0.4
$\{t_1, s_2\}$	0.17	0.2
$\{c_2\}$	0.33	0.4
$\{c_1\}$	0.33	0.4
$\{c_7\}$	0.5	0.6
$\{t_1\}$	0.33	0.4
$\{c_8\}$	0.83	1.0
$\{c_4, t_1, t_2\}$	0.17	0.2
$\{t_2, c_8\}$	0.17	0.2
$\{c_3, c_1\}$	0.17	0.2
$\{t_1, s_2, t_2\}$	0.17	0.2
$\{c_3, c_1, c_4\}$	0.17	0.2
$\{c_2, t_3, c_3, c_1, c_5\}$	0.17	0.2

TABLE V
VALUES OF m AND SALIENCE FOR SOME OF THE GROUPS IN TABLE III.

Group	m	Salience
$\{c_3\}$	0.5	0.5
$\{s_1\}$	0.5	0.5
$\{t_2\}$	1.0	1.0
$\{c_6\}$	0.75	0.75
$\{c_2, c_3\}$	0.25	0.25
$\{c_6, c_7\}$	0.25	0.25
$\{c_4, s_1, t_1\}$	0.5	0.5
$\{t_3, c_1\}$	0.25	0.25
$\{c_4\}$	0.5	0.5
$\{t_3\}$	0.5	0.5
$\{c_5\}$	0.75	0.75
$\{s_2\}$	1.0	1.0
$\{c_2, t_3\}$	0.25	0.25
$\{c_3, c_5\}$	0.25	0.25
$\{c_2, t_3, c_3, c_1\}$	0.25	0.25
$\{c_2\}$	0.5	0.5
$\{c_1\}$	0.5	0.5
$\{c_7\}$	0.75	0.75
$\{t_1\}$	0.5	0.5
$\{c_8\}$	1.0	1.0
$\{c_3, c_1\}$	0.25	0.25

groups (those containing the property gr in Γ) using Figure 2: the property "to include a group of two white objects" is true for the group of objects in the right of the image, but does not hold for the group of objects in the left, since the white objects in the latter do not form a group on the basis of distance. On the contrary, the property "to include a set of two white objects" holds for both groups.

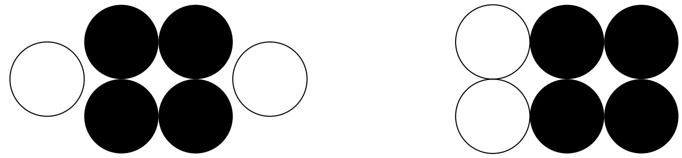

Fig. 2. Set vs. group example.

B. Expressions and referential success

As we mentioned before, the most usual kind of referring expression is that of a conjunction of properties. Such conjunctions can be univocally represented by a subset of properties $re \subseteq \mathcal{P}$ where, in the setting of this paper,

$$\mathcal{P} = Card \cup SP^P \cup P^{\subset}. \tag{14}$$

Given a referring expression $re \subseteq \mathcal{P}$ we can define the degree to which a set $g \in \mathcal{G}$ is a distinguishable group of objects satisfying all the properties in re as

$$re(g) = \min_{\gamma \in re} \gamma(g) \tag{15}$$

where $re(g) \in [0, 1]$.

Note that, given a referring expression re, $re(g)$ does not represent the degree to which re is a valid referring expression for g (in REG terminology, the referential success of re with respect to g) even when $re(g) = 1$. This is due to the fact that a referring expression is required not only to be true for the target, but also to distinguish it from the rest of potential targets, usually called *distractors*. For example, if there exists a $g' \neq g$ such that $re(g) = re(g') = 1$, the referential success of re for both g and g' is expected to be 0.

In [12]–[17], [21] we have studied the notion of referential success of referring expressions in a fuzzy framework, including axioms and different proposals. In this paper we shall employ the referential success measure we introduced in [13]. Let

$$\mathcal{G}_{re}(g) = re(g) \tag{16}$$

be a fuzzy set defined over $\mathcal{G}$ on the basis of re and let $\mathcal{G}^*_{re}$ be the normalization of this set performed by dividing by its maximum membership degree, assuming that it is not 0. Let us consider $\{a_1, \ldots, a_q\}$ as the set of membership degrees in $\mathcal{G}^*_{re}$ ranked in non-increasing order (that is, $a_i \geq a_j$ for each $1 \leq i < j \leq q$). Then, the referential success of re for g is defined as

$$RS(re, g) = \begin{cases} a_1(a_1 - a_2) & \text{when } condmax \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

where $condmax$ holds when $re(g) = \max_{g' \in \mathcal{G}} re(g') > 0$.

As an example, let us consider the image in Figure 3, which includes changes in gray tones for several objects with respect to Figure 1. The accomplishment degree of the property "to be dark" for the objects are shown in Table VI.

TABLE VI
DEGREES OF "DARK" FOR OBJECTS IN FIGURE 3.

c_1	c_2	c_3	c_4	c_5	c_6	c_7
1	0.9	1	0.6	0	0	0
c_8	t_1	t_2	t_3	s_1	s_2	
1	0.75	1	0.95	0.7	1	

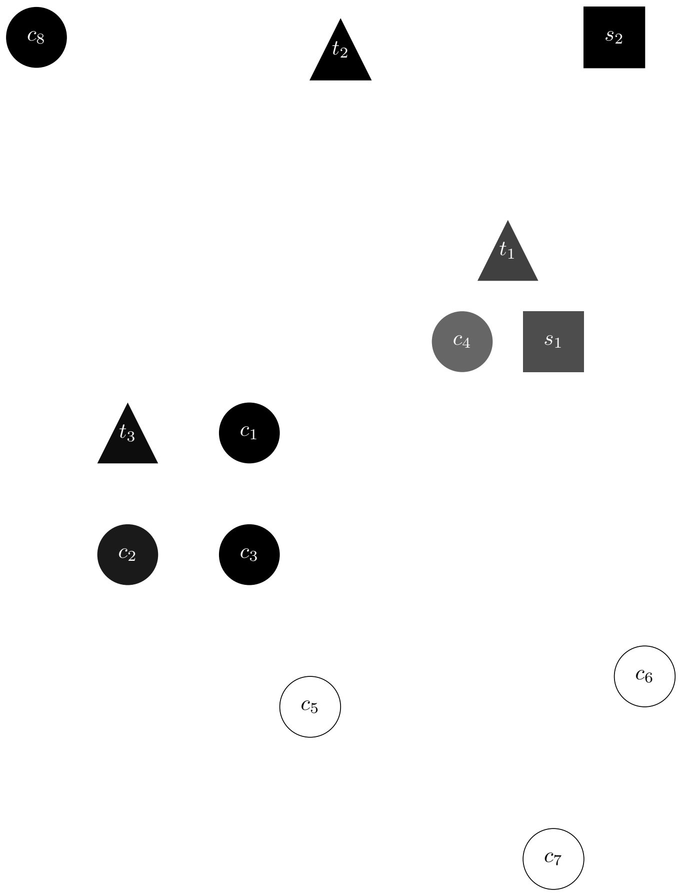

Fig. 3. A second scene example.

Table VII shows the accomplishment degree of the expression "to be a group of dark objects" (re_1). As can be seen, the expression produces high degrees in many of the considered groups and only obtains a referential success different from zero for group $\{c_3, c_1\}$ (whose $RS(re_1, g)$ is 0.05). In contrast, if we consider the expression "to be a group of three objects" (re_2), we obtain the results in table VIII. In this case, cardinality makes that the only referentiable group is $\{c_4, s_1, t_1\}$ with a maximum referential success.

The situation if we consider groups of Table II is rather different. There are more groups of three objects where re_2 expression fully holds, and its effect to distinguish $\{c_4, s_1, t_1\}$ disappears (RS(re_2,g)=0) for all g in Table II. Even if we add the property "dark" to re_2, the obtained referring expression re_3 "to be a group of three dark objects" is of little help. Table IX shows the obtained results. The expression only produces a low RS(re_2,g) of 0.75 $\times$ 0.05 = 0.0375 for the group $\{t_1, s_2, t_2\}$.

TABLE VII
VALUES OF $re_1(g)$ FOR GROUPS IN TABLE III.

$\{c_2, t_3, c_3, c_1\}$	0.9
$\{c_3, c_5\}$	0
$\{c_4, s_1, t_1\}$	0.6
$\{c_6, c_7\}$	0
$\{c_2, t_3\}$	0.9
$\{c_2, c_3\}$	0.9
$\{t_3, c_1\}$	0.95
$\{c_3, c_1\}$	1

TABLE VIII
VALUES OF $re_2(g)$ FOR GROUPS IN TABLE III.

$\{c_2, t_3, c_3, c_1\}$	0
$\{c_3, c_5\}$	0
$\{c_4, s_1, t_1\}$	1.0
$\{c_6, c_7\}$	0
$\{c_2, t_3\}$	0
$\{c_2, c_3\}$	0
$\{t_3, c_1\}$	0
$\{c_3, c_1\}$	0

In this case, the use of alternative expressions like "to be a group of three objects that includes a square and a circle" is neccesary to refer to $\{c_4, s_1, t_1\}$, with RS = 1.

C. Referring expression generation

The REG problem in the setting of groups of objects in images can be stated as follows: given a target group $g \in \mathcal{G}$, find a referring expression $re \subset \mathcal{P} = Card \cup SP^P \cup P^\subset$ with the highest possible referential success, given by Eq. (17).

This problem is known to be highly complex computationally, and it is approached by means of heuristic search techniques like Greedy algorithms. The latter algorithms choose one property in every step until a good enough referring expression is found, or all the properties have been considered. This procedure is linear in the number of properties, avoiding the full exploration of the inclusion lattice of subsets of properties. The different existing algorithms differ in the heuristic employed for choosing the most promising property in every step. For instance, the *discriminatory power* heuristic chooses the property that discards more distractors, whilst the Incremental Algorithm (IA) uses as heuristic a predefined ranking of the properties.

For the case of the properties we have introduced in the previous section, we suggest the following heuristics:

- Heuristics like the one used by the Incremental Algorithm can be employed in order to induce a ranking of the inclusion properties in $P^\subset$ corresponding to subgroups of properties, as follows:
  - $p^{g,\Gamma}$ is preferred to $p^{g',\Gamma'}$ when $sal(g) > sal(g')$, and
  - $p^{g,\Gamma}$ is preferred to $p^{g,\Gamma'}$ when $\Gamma < \Gamma'$ in the lexicographic order induced by the IA in the set of properties $SP^P \cup Card$.
- The discriminatory power heuristic can be directly employed by adding in each step the property that provides a larger increase of the referential success.

TABLE IX
VALUES OF $re_3(g)$ FOR GROUPS IN TABLE II.

$\{c_4, s_1, t_1\}$	0.6
$\{c_3, c_1, c_4\}$	0.6
$\{c_4, t_1, t_2\}$	0.6
$\{s_1, t_1, s_2\}$	0.7
$\{t_1, s_2, t_2\}$	0.75

As we have seen, the measure of salience can be used as a heuristic to rank properties in a REG approach in the setting of object groups based on the Incremental Algorithm. However, this measure can be useful for other purposes, particularly when choosing the target groups to be described in problems in which a whole description of the image has to be carried out, or when we are asked to give some brief but relevant information regarding the image.

IV. CONCLUSIONS AND FUTURE WORK

We have provided a preliminary approach to the two main tasks related to referring to groups of objects in images in a fuzzy setting: group identification and referring expression generation. Note that our approaches to these two tasks are independent, that is, our approach to referring expression generation can be applied to groups obtained by means of any other procedure; similarly, other approaches to referring expression generation can be applied to the groups we identify.

Regarding group identification, our approach allows to obtain crisp overlapping groups of objects that are coherent from the point of view of distance, and a measure of salience that is related to the difficulty of distinguishing a group in the image when our focus is restricted by a maximum distance d_m. Several important challenges remain to be afforded, since the identification of groups is a particularly complex problem that is affected by factors other than distance; however, we think that the use of fuzzy techniques, as well as those based on representations by levels, have much to contribute in this respect. In this sense, the salience measure proposed in this paper opens a way to explore, like for instance the study of the variation of the salience for different values of d_m, and its potential role to define a fuzzy concept of group.

Further future work will be related to the implementation of efficient algorithms for REG in this setting, following the guidelines we have introduced in the previous section, and its application in real problems. Quality models to assess interpretability and relevance of the resulting expressions will be also an object of future approximations.

REFERENCES

[1] E. Krahmer and K. Van Deemter, "Computational generation of referring expressions: A survey," *Computational Linguistics*, vol. 38, no. 1, pp. 173–218, 2012.

[2] M. Croitoru and K. van Deemter, "An inferential approach to the generation of referring expressions," in *Conceptual Structures: Knowledge Architectures for Smart Applications, 15th International Conference on Conceptual Structures, ICCS 2007, Sheffield, UK, July 22-27, 2007, Proceedings*, 2007, pp. 126–139.

[3] R. Fernández, "Salience and feature variability in definite descriptions with positive-form vague adjectives," in *Proceedings of the Workshop on the Production of Referring Expressions (PRE-CogSci 2009): Bridging the gap between computational and empirical approaches to reference*, 2009.

[4] S. Pezzelle and R. Fernández, "Is the red square big? malevic: Modeling adjectives leveraging visual contexts," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, 2019, pp. 2865–2876.

[5] E. Reiter and R. Dale, *Building natural language generation systems*. Cambridge university press, 2000.

[6] N. Marín and D. Sánchez, "On generating linguistic descriptions of time series," *Fuzzy Sets and Systems*, vol. 285, pp. 6–30, 2016.

[7] A. Ramos-Soto, A. Bugarín, and S. Barro, "On the role of linguistic descriptions of data in the building of natural language generation systems," *Fuzzy Sets and Systems*, vol. 285, pp. 31–51, 2016.

[8] K. van Deemter, *Computational Models of Referring: A Study in Cognitive Science*. Cambridge, MA: MIT Press, 2016.

[9] M. de Calmès, D. Dubois, E. Hüllermeier, H. Prade, and F. Sèdes, "Flexibility and fuzzy case-based evaluation in querying: An illustration in an experimental setting," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 11, no. 1, pp. 43–66, 2003.

[10] M.-J. Lesot, L. Mouillet, and B. Bouchon-Meunier, "Fuzzy prototypes based on typicality degrees," in *Computational Intelligence, Theory and Applications*, B. Reusch, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 125–138.

[11] A. Moreau, O. Pivert, and G. Smits, "A fuzzy approach to the characterization of database query answers," in *Information Processing and Management of Uncertainty in Knowledge-Based Systems - 16th International Conference, IPMU 2016, Eindhoven, The Netherlands, June 20-24, 2016, Proceedings, Part II*, ser. Communications in Computer and Information Science, J. P. Carvalho, M. Lesot, U. Kaymak, S. M. Vieira, B. Bouchon-Meunier, and R. R. Yager, Eds., vol. 611. Springer, 2016, pp. 329–340.

[12] A. Gatt, N. Marín, F. Portet, and D. Sánchez, "The role of graduality for referring expression generation in visual scenes," in *Proc. IPMU 2016, Volume 610 Communications in Computer and Information Science series*, J. P. Carvalho, M.-J. Lesot, U. Kaymak, S. Vieira, B. Bouchon-Meunier, and R. R. Yager, Eds. Springer, 2016, pp. 191–203.

[13] N. Marín, G. Rivas-Gervilla, and D. Sánchez, "A measure of referential success based on alpha-cuts," in *International Conference on Scalable Uncertainty Management*. Springer, 2016, pp. 345–351.

[14] ——, "Using specificity to measure referential success in referring expressions with fuzzy properties," in *Fuzzy Systems (FUZZ-IEEE), 2016 IEEE International Conference on*. IEEE, 2016, pp. 563–570.

[15] N. Marín, G. Rivas-Gervilla, D. Sánchez, and R. R. Yager, "Specificity measures and referential success," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 859–868, 2018.

[16] A. Gatt, N. Marín, G. Rivas-Gervilla, and D. Sánchez, "Specificity measures and reference," in *Proceedings of the 11th International Conference on Natural Language Generation, Tilburg University, The Netherlands, November 5-8, 2018*, 2018, pp. 492–502.

[17] N. Marín, G. Rivas-Gervilla, and D. Sánchez, "Referring under uncertainty," in *2019 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2019, New Orleans, LA, USA, June 23-26, 2019*, 2019, pp. 1–6.

[18] D. Dubois and D. Sánchez, "Fuzzy clustering based on coverings," in *Towards Advanced Data Analysis by Combining Soft Computing and Statistics*, ser. Studies in Fuzziness and Soft Computing, C. Borgelt, M. Gil, J. Sousa, and M. Verleysen, Eds. Springer, 2013, pp. 319–330.

[19] D. Sánchez, M. Delgado, M. Vila, and J. Chamorro-Martínez, "On a non-nested level-based representation of fuzziness," *Fuzzy Sets and Systems*, vol. 192, no. 1, pp. 159–175, 2012.

[20] N. Marín, G. Rivas-Gervilla, and D. Sánchez, "An approximation to context-aware size modeling for referring expression generation," in *2018 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2018, Rio de Janeiro, Brazil, July 8-13, 2018*, 2018, pp. 1–8.

[21] N. Marín, G. Rivas-Gervilla, D. Sánchez, and R. R. Yager, "Specificity measures based on fuzzy set similarity," *Fuzzy Sets and Systems*, In press, https://doi.org/10.1016/j.fss.2019.06.001.