Describing Images Via Linguistic Features and Hierarchical Segmentation

R. Castillo-Ortega, J. Chamorro-Martínez, N. Marín, *Member, IEEE*, D. Sánchez, and J.M. Soto-Hidalgo

***Abstract*— In this paper we introduce a preliminary proposal for linguistic description of images. The approach is based on i) a hierarchical fuzzy segmentation of the image, ii) a collection of linguistic features describing each region, and iii) fuzzy spatial relations and locations. The procedure is independent from the way these elements have been obtained, and provides a description with the characteristics of a summary, i.e., a *brief* and *accurate* description of the *whole* image. As another characteristic of summaries, the method can be guided in the description by the user's preferences and interest. Remarkably, we are able to provide a description containing sentences about disjoint regions appearing in different levels of detail.**

I. Introduction

Filling the semantic gap between computer representation of digital images and human description is one of the main topics in the area of Computer Vision, known as the *semantic gap* [1]. The classical application is that of automatic image annotation, i.e., the automatic identification of semantic concepts appearing in an image. This is particularly relevant for retrieving images on the basis of linguistic queries, for which approaches based on specifying queries as images or sketches are not well suited. The result of this procedure is a set of terms representing concepts that "appear" in the image.

One of the main difficulties of this task is the fact that the "appearance" of a given term in an image, as identified by a human subject, is in fact a matter of association between the visual features perceived in the image and the internal representation of the concept represented by the term in the subject's mind. While visual features can be extracted from the image, modeling our internal representation of concepts and the association between both is much more difficult. Current approaches are based mainly on building classifiers on the basis of a collection of samples, using any of the techniques available in machine learning. This is currently a very active field of research.

However, there are richer types of linguistic descriptions with interesting applications. In this work we are concerned with obtaining a linguistic description of an image using linguistic descriptions of regions obtained by a segmentation process, and terms related to the relative position of different regions. Contrary to the semantic annotation problem, the expected result in the case of linguistic descriptions is a small set of sentences, built on the basis of a certain subset of natural language. This problem has been paid little attention in the literature, though some related work can be found like for example [2], [3], [4], [5], [6]. There are many potential applications of this technology such as fast transmission of preliminary visual information, automatic sports commentators in video games [7], description of diseases visible in X-Ray images [4], [8], or generating descriptions of images for the visually impaired, envisioned in [2].

The main contribution of this work is related to using a hierarchical fuzzy segmentation of the image as the guide to obtain the description. Part of this problem is solved by adapting a time series summarization technique developed by some of the authors in a previous work; the approach intends to provide an *accurate* and at the same time *brief* description of a collection of fuzzy regions that form a partition of the *whole* image. We shall illustrate the proposal by using a description of regions in terms of color concepts coming from a recent proposal for modeling the semantics of colors by means of *fuzzy color spaces* proposed in [9], with the ability of adapting the fuzzy color space to the characteristics of the image and/or the user's perception of color. However, we shall discuss how this preliminary work can (and will be) easily extended to other image features and high-level concepts.

The paper is organized as follows. In Section II we introduce important notions related to fuzzy segmentation and fuzzy hierarchical segmentation that are on the basis of our approach, as well as important notions of absolute locations and spatial relationships. The latter are employed in order to provide an interpretation of images as an edge-labeling of a graph whose vertices are fuzzy regions of any level of the hierarchy. In Section III we introduce our approach to linguistic description based on color terms. We first describe our approach to the design of fuzzy color spaces; then we employ this tool in order to obtain a summary of color information in the image. Finally, a linguistic description of regions with absolute locations, spatial relationship, and dominant color is provided. We illustrate our approach with an example in Section IV. Finally, Section V contains our conclusions and some comments on future research.

II. Segmenting Images as Labeled Graphs

A. Hierarchical segmentation

The relevant regions in an image are usually obtained by using image segmentation techniques. Among the huge

Corresponding author: D. Sánchez

D. Sánchez, N. Marín, J. Chamorro-Martínez, and R. Castillo-Ortega are with the Department of Computer Science and Artificial Intelligence, University of Granada, C/ Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain. e-mail: {daniel,nicm,jesus,rita}@decsai.ugr.es.

J.M. Soto-Hidalgo is with the Department of Computer Architecture, Electronics and Electronic Technology, University of Córdoba, Spain. e-mail: jmsoto@uco.es

This work was supported by the Spanish Government (Science and Innovation Ministry) under project TIN2009-08296 and also by the Andalusian Government (Junta de Andalucía) under project P07-TIC03175.

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

amount of techniques available in the literature we consider specially appropriate to use *fuzzy segmentation techniques*, since boundaries of regions are fuzzy in most of the cases. Fuzzy segmentation obtains a collection of fuzzy regions (fuzzy subsets of connected pixels with similar features) that form a *fuzzy partition* in some sense of the pixels in the image [10], [11], [12], [13], [14], [15], [16].

A well known problem in image segmentation is that of finding the appropriate amount of segmentation detail for an specific application. A usual solution consists in using hierarchical segmentation. Hierarchical segmentation consists in obtaining an ordered set of nested segmentations, each segmentation being a level in the hierarchy, verifying that a region in one level is included in one region of the next level. Each level represents a different amount of detail in the description of the image. There are many approaches to hierarchical segmentation of images, like [17], [18], [19], [20]. The use of hierarchical segmentation in image processing and computer vision is very important in applications like image compression [21], [22], scene description and image parsing [23], knowledge discovery in images [24], and remote sensing [25], among many others [26], [27].

In [28], [29] an approach to obtain a fuzzy hierarchical segmentation on the basis of a fuzzy segmentation is proposed. This procedure can be applied to fuzzy segmentations obtained by any means; we usually apply it on a fuzzy segmentation obtained by using the method in [16].

The approach to image description we present in this paper is independent of the way the hierarchy has been obtained. Without losing generality, we shall assume a fuzzy hierarchy (that may be crisp as a particular case) characterized as follows: the segmentation of the image is hierarchically organized in n levels, namely, $L = L_1, ..., L_n$. Each level L_i has associated a fuzzy segmentation of the image in p_i regions $\{D_{i,1}, ..., D_{i,p_i}\}$. The membership functions for fuzzy regions are assumed to be normalized. We assume that each level contains a fuzzy partition of the pixels in the image, in the following sense: a set of labels $\{X_1, ..., X_r\}$ is a fuzzy partition on X iff:

1) $\bigcup_{i\in\{1,...,r\}} Support(X_i) = X$.
2) $\forall i, j \in \{1, \dots, r\}, i \neq j, Core(X_i) \cap Core(X_j) = \emptyset$.
3) $\forall i \in \{1, \dots, r\}\ \exists x \in X$ such that $X_i(x) = 1$, i.e., there is at least one object fully representative of X_i.

Condition 3 is always verified by definition of fuzzy region. Condition 1 is equivalent to $\forall x \in X\ \exists i \in \{1, \dots, r\}$ such that $X_i(x) > 0$. Conditions 2 and 3 imply $\widetilde{C}_i \not\subseteq \widetilde{C}_j\ \forall i \neq j$.

Additionally, considering the hierarchy, we add the following constraints:

1) $\forall i, j \in \{1, \dots, n\}$, $i < j$ implies $p_i > p_j$ (i.e, as we move upward in the hierarchy, the number of labels in the partition decreases).
2) $\forall i \in \{2, \dots, n\}, \forall j \in \{1, \dots, p_i\}, \forall k \in \{1, \dots, p_{i-1}\}$ $(D_{i,j} \subseteq D_{i-1,k})$ implies $(D_{i,j} = D_{i-1,k})$ (i.e., labels cannot generalize another label of an upper level).

B. Spatial Relationships

The spatial arrangement of objects provides key information for image description. Using the result of a fuzzy segmentation process, an image can be interpreted as a graph in which regions are the vertices, and two regions are connected when they share a boundary. An improvement of this representation is obtained when edges are labeled with the spatial relationship between the regions, yielding an edge-labeled graph. In order to obtain such labeled graphs from images it is necessary to solve two main problems:

1) To determine the possible spatial relationships between regions and to provide a suitable linguistic term for each one.
2) To provide a procedure for determining the spatial relationship that hold for a certain couple of regions.

There are many approaches for the first problem, that have been modeled as ontologies in some works. The RCC-8 model proposed in [30] is used for the ontology in [31], [32], and a fuzzy version (fuzzy RCC-8) proposed in [33] is employed in [32]. In [2] the proposal in [34] is employed. In [35] an ontology is built taking into account the spatial relationships proposed in [36] and [37]. Regarding the determination of spatial relationships, methods based on computing on the localization of couple of points are employed in [38], [39], [40], as pointed out in [41]; this latter work employs for the reverse problem (i.e., generating sketches from linguistic descriptions) a method based on the "histogram of forces" [42], [43]. Finally, approaches for defining complex spatial relationships in terms of simpler ones by using logic expressions are proposed (among others) in [44], using a three-valued logic, and [32] on the basis of fuzzy description logic. Remarkably, [44] discuss about spatial relations in hierarchical segmentations.

We consider in this work the fuzzy RCC-8 model as described in [32]. The spatial relationships in this model, together with the corresponding expressions in a fuzzy description logic as proposed in [32], are shown in Table I.

TABLE I

FUZZY RCC-8 SPATIAL RELATIONSHIPS.

Name	**Relation**	**RCC definition**
Disconnected	DC	$\neg C(a, b)$
Part	P	$\forall c.(C(c, a) \to C(c, b))$
Proper Part	PP	$P(a, b) \wedge \neg P(b, a)$
Equals	EQ	$P(a, b) \wedge P(b, a)$
Overlaps	O	$\exists c.(P(c, a) \wedge P(c, b))$
Discrete	DR	$\neg O(a, b)$
Partially Overlaps	PO	$O(a, b) \wedge \neg P(a, b) \wedge \neg P(b, a)$
Externally connected	EC	$C(a, b) \wedge \neg O(a, b)$
Non Tangential Part	NTP	$\forall c.(C(c, a) \to O(c, b))$
Tangential PP	TPP	$PP(a, b) \wedge \neg NTP(a, b)$
Non-Tangential PP	NTPP	$PP(a, b) \wedge NTP(a, b)$

Finally, let us remark that our approach to linguistic description of images is independent of the set of spatial relationships employed, and the way they are assessed in an image. In the following we shall assume that a set of spatial relationships $\mathcal{SR} = \{R_1, \dots, R_k\}$ is available.

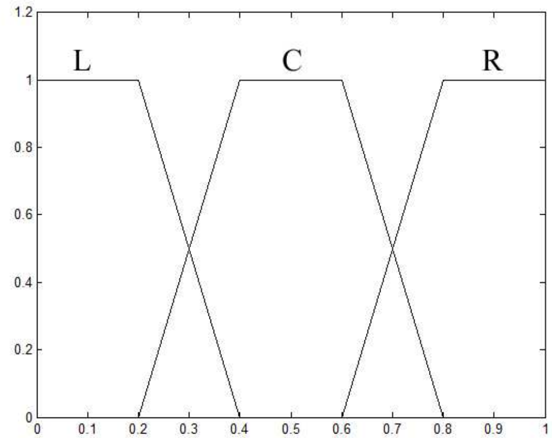

Fig. 1. Fuzzy horizontal position. L: left; C: Center; R:right.

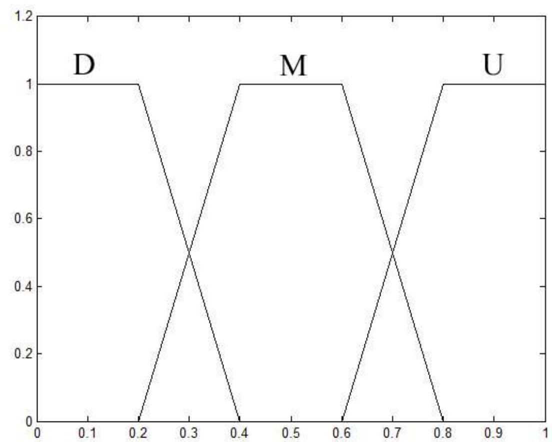

Fig. 2. Fuzzy vertical position. D:down; M: Middle; U:up.

C. Absolute Locations

Spatial relationships describe the locations of regions with respect to others. However, linguistic description requires to model the absolute location of regions in the image as well [31]. These absolute locations may be seen as relative with respect to the image's boundaries.

In this paper we propose to employ a fuzzy partition of the image for defining the absolute location of a region. The proposed partitions, defined on the domain of the percentages of the horizontal and vertical lengths, are shown in Figures 1 and 2, respectively. The cartesian product of both partitions using the minimum yield a fuzzy partition of the area of the image as shown in Figure 3. This partition can be refined by using more labels in both lengths if needed.

We determine the degree to which a given fuzzy region D is in a fuzzy absolute location L by evaluating a quantified sentence of the form Q_M of D are L, where $Q_M(x) = x$, using method GD [45]. This method verifies suitable properties for a fuzzy set inclusion measure. The evaluation

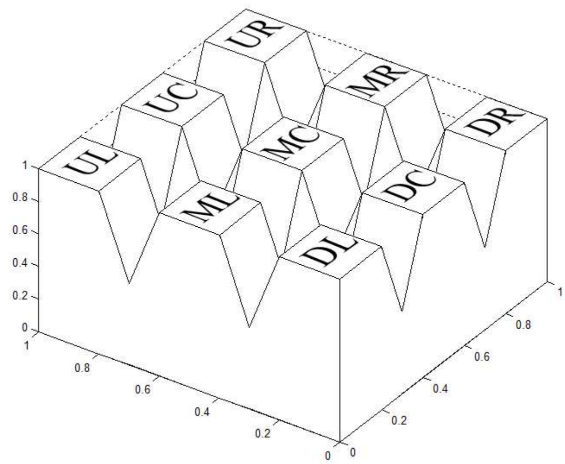

Fig. 3. Fuzzy absolute locations as the combination of fuzzy horizontal and vertical lengths.

of "Q of D are A" by means of GD is

$$GD_Q(A/D) = \sum_{\alpha_i \in \triangle(A/D)} (\alpha_i - \alpha_{i+1}) Q\left(\frac{|(A \cap D)_{\alpha_i}|}{|D_{\alpha_i}|}\right) \quad (1)$$

where $(A \cap D)(x) = \text{Min}(A(x), D(x))$, $\triangle(A/D) = \Lambda(A \cap D) \cup \Lambda(D)$, $\Lambda(D)$ being the level set of D, and $\triangle(A/D) = \{\alpha_1, ..., \alpha_p\}$ with $\alpha_i > \alpha_{i+1}$ for every $i \in \{1, ..., p\}$, $\alpha_1 = 1$ and $\alpha_{p+1} = 0$. The set D is assumed to be normalized. If not, D is normalized and the same normalization factor is applied to $A \cap D$ [1].

The absolute locations shown in Figure 3, or any finer-grained alternative, may be enriched by a hierarchical clustering of locations. The idea is that large regions may not be included in any of the locations because of their size; hence, considering larger, more imprecise locations as the union of locations is necessary. With this idea we can even provide a complete ontology of locations in which for example the union of labels DL, DC, and DR is simply called "Down", and the union of all the labels except MC is called "Perimeter". In order to determine the location in this ontology that best matches the location of a given fuzzy region, we shall consider a minimum accomplishment degree of the corresponding quantified sentence, and we shall look for sentences starting with the more precise absolute locations. This way we avoid to obtain large locations as much as possible.

D. Edge-Labeled Graph of Regions

Let $\mathcal{SR} = \{R_1, \ldots, R_k\}$ be a set of fuzzy spatial relations. For a certain fuzzy segmentation of an image, let $D_i = \{D_{i,j}$ such that $, j \in \{1, \ldots, p_i\}\}$ and let

[1] As the data set is finite, D is considered to be finite, and hence the number of relevant α-cuts is also finite.

$$D = \bigcup_{i \in \{1,\ldots,n\}} D_i \quad (2)$$

i.e., D is the set of all the fuzzy regions appearing in all the levels of the hierarchy. Then, we call Graph of Regions of the image to the directed graph $G = (D, E)$ where the vertices are the regions in D, and there is a directed edge between each two different regions $D_{i,j}$ and $D_{k,l}$ in D. The labeling of the graph consists in an assignment of integers in $\{1, \ldots, k\}$ to the edges, so that the edge linking regions $D_{i,j}$ and $D_{k,l}$ will be labeled with the integer z iff the fuzzy spatial relation that best matches the spatial relationship between both regions is R_z.

III. Our Approach to Linguistic Description of Images

We shall explain here our approach to linguistic description on the basis of color terms and region locations. This approach will be extended to other linguistic terms in forthcoming works. The approach works in two steps:

1) First, we obtain a summary of color information in the images by assigning the most representative color term to each region of a certain subset. The representativity is measured in terms of the amount of pixels that match that color, and a minimum threshold for this value given by the user in the form of a fuzzy quantifier and a minimum accomplishment degree. This guarantees that the summary is *accurate*. The subset is determined as to form a partition of the image with a minimal collection of fuzzy regions, hence looking for *brevity* of the summary and, at the same time, *covering the whole* image.
2) We build a collection of natural language sentences taking into account the summary of the previous step, and spatial information in the form of absolute locations and the spatial relationships in the edge-labeled graph.

In the next section we briefly describe our approach to modeling colors by formal definitions of the notions of fuzzy color and fuzzy color space. Our approach to summarization of the image in terms of colors is presented in Section III-B. Finally, we explain how we obtain the final linguistic description in Section III-C.

A. Linguistic Characterization of Color in Regions: Fuzzy Color Spaces

In computer science, color is usually represented as a triplet of real values in different ways, each one with different domains and semantics for the real values. Each such systems is called a *color space*. A well known example is the RGB color space, where the semantics of the three values defining a color in this case (integers in [0,255]) is the amount of red, green, and blue necessary to provide the color.

We humans are able to use a small number (up to 300) of colors in comparison with the expressive power of color spaces (several millions), and we use to express them by means of linguistic terms. For example, we do not employ the numerical triplet $[255, 0, 0]$ in our discourse, but we say *red*; furthermore, there is no biunivocal link between linguistic terms and colors in a color space, but each linguistic term corresponds to a subset of colors. Unfortunately, the boundaries of such set representations are imprecise, subjective, and depend on the application domain and cultural issues.

The lack of a clear correspondence between color spaces and linguistic terms is a clear example of what is known as the "semantic gap", and constitutes an important problem for applications required to support natural language. In order to manage the imprecision in color description, we introduced in [9] the following definitions:

Definition 3.1: A fuzzy color $\widetilde{C}$ is a linguistic label whose semantics is represented in a color space XYZ by a normalized fuzzy subset of $D_X \times D_Y \times D_Z$.

Definition 3.2: A fuzzy color space $\widetilde{XYZ}$ is a set of XYZ fuzzy colors that define a fuzzy partition of $D_X \times D_Y \times D_Z$.

In this last definition, the notion of fuzzy partition we employ is that we introduced in Section II-A.

In the same work we present a proposal for building a customized fuzzy color space, taking as starting point a set of crisp colors $R = \{\mathbf{r}_1, \ldots, \mathbf{r}_m\}$ fully representative of the fuzzy colors we want to obtain. For each $\mathbf{r}_i$ we will obtain an atomic fuzzy color $\widetilde{C}_i$ on the basis of a partition of the RGB color space, using the euclidean distance d for obtaining the membership function. A more detailed description can be found in [9]. Let us remark that the obtained colors verify the following properties:

1) The fuzzy sets obtained are normalized and convex.
2) The set of fuzzy sets obtained form a fuzzy color space since they are a fuzzy partition in the sense indicated in Section II-A.
3) $\widetilde{C}_i(\mathbf{c}) > 0.5$ iff $d(\mathbf{r}_i, \mathbf{c}) < d(\mathbf{r}_j, \mathbf{c})\ \forall i \neq j$.
4) If a crisp color $\mathbf{c}$ is equidistant from two representatives $\mathbf{r}_i$ and $\mathbf{r}_j$ then $\widetilde{C}_i(\mathbf{c}) = \widetilde{C}_j(\mathbf{c}) = 0.5$

Using our methodology, we have developed fuzzy color spaces using color names provided by the well-known ISCC-NBS system [46], [47]. This system is based on the pioneering work of Berlin and Kay [48] about color naming and has been tested with humans on a task of color description, hence it is suitable for our purpose. ISCC-NBS provides several color sets in the form of sets of pairs (linguistic term, crisp color), and we have developed three fuzzy color spaces using the sets *Basic*, *Extended*, and *Complete*, defined as follows:

- **Basic Set**: 13 color names corresponding to ten basic color terms (pink, yellow, red, orange, brown, olive, green, blue, violet, purple), and 3 achromatic ones (white, grey, and black).
- **Extended Set**: 31 color names corresponding to those of the basic set and some combination of them formed by adding the $-ish$ suffix (Brownish Orange, Purplish Blue among others).
- **Complete Set**: 267 color names obtained from the extended set by adding five tone modifiers for lightness (very light, light, medium, dark and very dark) and four

adjectives for saturation (grayish, moderate, strong and vivid). Also, three additional terms substitute certain lightness-saturation combination (pale for light grayish, brilliant for light strong and deep for dark strong). Theses color names are represented by the Universal Color Language, Level 3 in the ISCC-NBS system.

Let us remark that the choice of one fuzzy color space or another greatly influences the final linguistic description. For example, among the three previous spaces, the Basic one allows us to obtain a more brief description, since there are less fuzzy colors and there are more crisp colors matching each fuzzy one, hence it will be easier to find large regions containing mostly pixels with a single fuzzy color. On the contrary, the Complete fuzzy color space provides descriptions with more detailed and precise colors. The extended space constitutes a compromise between them. Let us also remark that these are only three examples, and that our method for linguistic summarization can be employed on the basis of any color space. Our methodology hence allows the user to specify the color term set she wants to use in the description and, indirectly, will affect the length and precision as we have just seen.

B. Summarizing the Image Using Fuzzy Colors and Absolute Locations

Our summarization in terms of color will be based on adapting the technique proposed in [49] for the linguistic summarization of time series. Notice that both fuzzy regions and fuzzy colors define fuzzy subsets of pixels in the image. We are then interested in a color summary that take the form of a collection of quantified sentences describing the most frequent colors in as few regions as necessary to describe the image. The approach in [49] will deliver a collection of sentences of the form "Q of $D_{i,j}$ are A" where:

- $D_{i,j}$ is a label member of a certain level i of the hierarchy associated to the fuzzy segmentation of the image.
- A is a fuzzy color from the fuzzy color space chosen by the user.

The user must provide a collection of quantifiers defining the kind of fuzzy quantities and percentages he/she is interested in. This can be defined by choosing among a collection of predefined quantifiers. In [49] we consider that the user provides a *totally ordered* subset $\{Q_1, ..., Q_{qmax}\}$ of a coherent family of quantifiers $\mathcal{Q}$ [50] to be used in the summarization process.

Definition 3.3 (Coherent family of quantifiers, [50]): Let $\mathcal{Q}$=$\{Q_1$, ..., $Q_l\}$ be a linguistic quantifier set, we shall say it is coherent if it verifies that:

- The membership functions of $\mathcal{Q}$ elements are non-decreasing functions.
- A partial order relation $\succeq$ is defined in $\mathcal{Q}$. It has as its maximal element $Q_1 = \exists$ and as its minimal one $Q_l = \forall$. Furthermore $\forall Q_i, Q_j \in \mathcal{Q}, Q_i \subseteq Q_j \Rightarrow Q_j \succeq Q_i$.
- The membership function of the quantifier $\exists$ is given by $Q_1(x) = 1$ if $x \neq 0$ and $Q_1(0) = 0$, whereas the membership functions of $\forall$ will be $Q_l(x) = 0$ if $x \neq 1$ and $Q_1(1) = 1$.

1) $ToSummarize \leftarrow L_n$;
2) $Summary \leftarrow \emptyset$; $Summarized \leftarrow \emptyset$;
3) While $ToSummarize \neq \emptyset$
    a) Take $D_{i,j} \in ToSummarize$
    b) $ToSummarize \leftarrow ToSummarize \backslash \{D_{i,j}\}$
    c) $p \leftarrow qmax$; $covered \leftarrow false$;
    d) While $p \geq Qbound_i$ and $not\ covered$
        i) $k \leftarrow 1$;
        ii) While $k \leq Gbound_i$ and $not\ covered$
            A) Let $A \leftarrow argmax_{B \in C_k} GD_{Q_p}(B/D_{i,j})$
            B) If $GD_{Q_p}(B/D_{i,j}) \geq \tau$ then $Summary \leftarrow Summary \cup \{Q_p$ of $D_{i,j}$ are $A\}$; $Summarized \leftarrow Summarized \cup (D_{i,j})$ $covered \leftarrow true$;
            C) $k \leftarrow k + 1$;
        iii) $p \leftarrow p - 1$
    e) If $not\ covered$ and $i > 1$ then $ToSummarize \leftarrow ToSummarize \cup ch(D_{i,j})$.
    f) else if $i = 1$ then $Summary \leftarrow Summary \cup \{D_{i,j}$ is highly variable$\}$

Fig. 4. Algorithm to obtain linguistic summaries on the basis of color.

In addition, the user will provide a threshold τ for the minimum accomplishment degree he/she wishes for the quantified sentences comprising the summaries. We shall evaluate this degree by using again the method GD introduced in Section II-C, Eq. 1.

The requirements for this collection of quantified sentences, according to the intuitive idea of summary, are the following:

- The accomplishment degree of every sentence must be greater or equal than τ, i.e., the information provided by every sentence must hold in the image to a high (τ) degree (accuracy).
- The set of quantified sentences must be as small as possible (brevity).
- The regions $D_{i,j}$ in the sentences of the summary must comprise a fuzzy partition of the image (coverage). We shall denote this set of regions by $D' \subseteq D$.

In [49] we proposed the algorithm given by 4 as an implementations to meet the above mentioned goals[2]. In this algorithm, $ch(D_{i,j})$ is defined as follows: $ch(D_{1,j}) = \emptyset$ for all j. Otherwise, $ch(D_{i,j}) = \{D_{i-1,k}, k \in \{1..p_{i-1}\} | D_{i-1,k} \cap D_{i,j} \neq \emptyset$ and $\neg \exists D \in ToSummarize \cup Summarized$, $(D_{i-1,k} \cap D_{i,j}) \subseteq D\}$, and $C_k = \{\cup_{E_h \in F} E_h \mid F \subseteq E, |F| = k\}$.

In order to look for brevity, we start from the fuzzy regions in the top level of the hierarchy. Each level has its own quantifier bound ($Qbound_i$) and grouping bound ($Gbound_i$)

[2]Notice that many solutions may fulfill these conditions, and we do not look for the optimum.

1) Arrange the fuzzy regions in D' in a total order. Suppose $D' = \{D^1, \ldots, D^u\}$ with $D^i \prec D^{i+1}$ $\forall i$.
2) Generate a linguistic description of D^1
3) $i \leftarrow 2$
4) While $(i \leq u)$
    a) Include the spatial relationship between D^{i-1} and D^i
    b) Generate a linguistic description of D^i
    c) $i \leftarrow i + 1$

Fig. 5. Algorithm to obtain the final linguistic description.

that, respectively, indicate the less strict quantifier to be considered and the maximum number of color labels to be aggregated in a sentence at this level of the time domain; this means that, when $Gbound_i > 1$, we allow the algorithm to obtain color descriptions as disjunctions of colors like for example *red or orange*. In principle, we shall set $Gbound_i = 1$ $\forall i$. The set $ToSummarize$ is the collection of regions in the image for which a quantified sentence is missing. If it is possible to obtain an accomplishment degree greater than τ for a certain area using a quantifier Q and a single label, the procedure obtains a summary for that region. If it is not possible, the procedure tries with the union of different subsets of color labels: couples, trios, quartets, etc, until we obtain an accomplishment degree greater than τ. The size of the subset is given by k being $Gbound$ its maximum value. When a summary is found in a certain region we say that the region is $covered$. If all the groups were tried without success, the algorithm repeats the grouping process again, but with a less strict quantifier, until $Qbound_i$ is reached. If there is no result found for a given region $D_{i,j}$, we try to obtain such sentences with the *corresponding children* $ch(D_{i,j})$ in the next level. If $ch(D_{i,j}) = \emptyset$, then a sentence indicating the observed color variability is added to the summary ($D_{i,j}$ *is highly variable in color*). The final set of linguistically quantified sentences comprising the summary is $Summary$.

Our approach provides a fuzzy partition of the image in the form of a collection of fuzzy regions D', and associates to each fuzzy region in D' a fuzzy color and a linguistic quantifier expressing the amount of pixels in the region that match the fuzzy color. An important property of D' is that the fuzzy regions that form this fuzzy partition may not pertain to the same level of the hierarchy. This is a solution to the important problem of multiple-scale segmentation pointed out in [27].

C. Generating the Linguistic Description

Once the color summary of the image is available, we can use this information to generate the linguistic description. The generic procedure we propose is detailed in Figure 5.

In this very general algorithm there are open issues. First of all, how to arrange the fuzzy regions for the description. Some authors, like for example [51], propose to use preference measures reflecting the interest of the user. We consider different possibilities of arrangement in terms of position, size, color, or any combination of them. In this paper we shall assume that the fuzzy regions are arranged by size. Second, the linguistic description to be generated can incorporate different elements. In this paper we shall employ linguistic terms relative to color and absolute location. Finally, the spatial relationship between regions is employed for linking the different descriptions of regions in a kind of scene navigation.

As pointed out in [49], a further summary may be obtained by joining together sentences that share a common descriptor, either about color, location, etc. We shall employ this choice in the next section.

IV. Example

We shall illustrate our approach with a simple example on the image in Figure 6. In this figure, the blue squares indicate the seeds employed in a region-growing fuzzy segmentation algorithm. We can see that some of these seeds have been placed in the same regions, hence we will obtain an oversegmentation of the image. Following the segmentation algorithm proposed in [16] we shall obtain two overlapping regions; obviously, we are interested in describing only one of them. As we shall see, our algorithm will solve this point.

Starting from the mentioned segmentation algorithm, a fuzzy hierarchical segmentation is obtained by using the proposal in [28], [29]. The hierarchical segmentation is shown in Figure 7. The hierarchy consists of eight levels $L_1, \ldots, L_8$, described in eight columns. In each column the membership function of each of the fuzzy regions comprising the segmentation at that level are shown as images in which the color white means membership degree 1 to the fuzzy region, and black means 0. Intermediate gray colors correspond to intermediate degrees. The inclusion of fuzzy regions in one level with respect to regions in other levels are indicated. The union of fuzzy regions is obtained via the maximum.

The first level L_1 of the hierarchy correspond to the fuzzy segmentation obtained by the algorithm proposed in [16], and consists of eight fuzzy regions corresponding to the eight seeds in Figure 6. We shall number the fuzzy regions in each level from top to bottom. For instance, in level L_2 the fuzzy region $D_{2,1}$ corresponds to the union of fuzzy regions $D_{1,1}$ and $D_{1,2}$ of level L_1.

We have considered a single trapezoidal quantifier $Q = (0, 0.7, 0.9, 1)\}$ that we have called *most of* and a threshold $\tau = 0.7$ as parameters, and we have not considered grouping colors. Other elements employed have been the fuzzy color space we have designed on the basis of the Basic set of colors of the ISCC-NBS system, the absolute locations defined in Figure 3, and the spatial relationships in Table I.

In the summary phase, we apply the algorithm in Figure 4. The set *ToSummarize* is initialized to $\{D_{8,1}\}$, as this is the last level L_8 in the hierarchy. The algorithm is unable to find a good summary for the fuzzy region $D_{8,1}$ in level L_8 (as

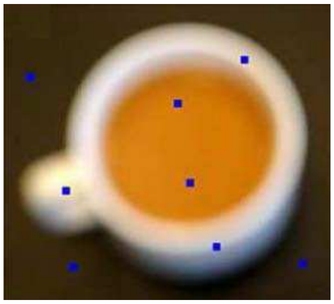

Fig. 6. Example image.

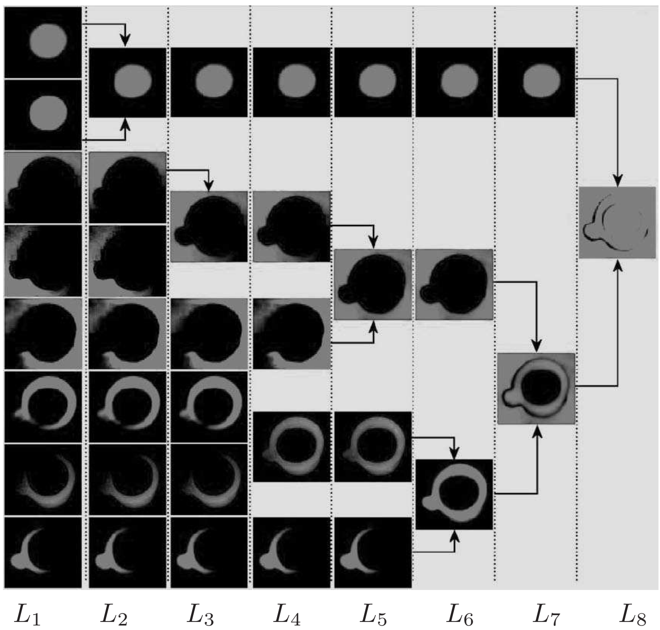

Fig. 7. Hierarchical segmentation of the image in Figure 6.

expected, since it covers the whole image, and this image is not homogeneous in color), and proceeds eliminating this fuzzy region from the set *ToSummarize* and adding to the same set the children of $D_{8,1}$ in level L_7, $ch(D_{8,1}) = \{D_{7,1}, D_{7,2}\}$. For $D_{7,1}$, the algorithm is successful in finding a fuzzy color such that the evaluation of the quantified sentence is above the threshold τ; the resulting sentence is *Most of* $D_{7,1}$ *are Orange*, and $D_{7,1}$ is eliminated from *ToSummarize*. However, the algorithm fails to find a good summary for the fuzzy region $D_{7,2}$, so this is removed from *ToSummarize* and the corresponding children in level L_6, $ch(D_{7,2}) = \{D_{6,2}, D_{6,3}\}$ are added to the same set. For these two regions, the algorithm provides sentences with degree above τ, *Most of* $D_{6,3}$ *are White*, and *Most of* $D_{6,2}$ *are Olive*, eliminating both $D_{6,2}$ and $D_{6,3}$ from *ToSummarize*. Since *ToSummarize*$= \emptyset$, the procedure stops here, since that means that the whole image has been covered by the description.

Additionally, the absolute location of these regions is determined. In this case the regions are rather large in proportion to the size of the image. Employing a low threshold for the fulfilment of the quantified sentences for inclusion, we obtain that the location of $D_{7,1}$ and $D_{6,3}$ is CC (*Medium-Center*), though for $D_{6,3}$ the degree obtained is almost the same for *Perimeter*; for $D_{6,2}$ the position is *Perimeter*. The spatial relationships that hold with larger degree between regions are $NTPP(D_{7,1}, D_{6,3})$ and $NTPP(D_{6,3}, D_{6,2})$.

For the final generation of the linguistic description, the regions are arranged in the order $D_{6,2}, D_{6,3}, D_{7,1}$. The final linguistic description is generated by describing each of these regions in this order in terms of its position and color. For the first region, the absolute position and color are given. Then, the relative position between the next region and the previous one, and the color of this region, until the last region is described. The result is

> *There is a region in the Perimeter of the image with Most of color Olive; this has a non-tangential proper part in the Medium-Center of the image with Most of color White; this has a non-tangential proper part in the Medium-Center of the image with Most of color Orange.*

V. Conclusions

We have proposed a methodology for the linguistic description of images on the basis of color information. The starting point is a fuzzy hierarchical segmentation. Since different people would provide different summaries of the same information, depending on their interest and/or previous knowledge, we have considered many degrees of freedom in our methodology. These are given by the definition of fuzzy absolute locations, fuzzy spatial relationships, fuzzy color spaces, quantifiers, accomplishment degrees, grouping bounds, and criteria for ranking the fuzzy regions that form the partition of the image provided by our summarizing algorithm. The method is able to solve the problems of oversegmentation and multiple-scale segmentation of the images, choosing regions in different levels of the hierarchical segmentation on the basis of their suitability. There are still many open issues for future research, some of which have been pointed out in previous sections. In particular, we shall consider other characteristics of fuzzy regions like texture and shape.

Acknowledgment

The authors are very grateful to Pedro Martínez-Jiménez for his valuable help in the preparation of this paper.

References

[1] A. Smeulders, M.Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.

[2] D. Roy, "Learning visually grounded words and syntax for a scene description task," vol. 16, no. 3, pp. 353–385, 2002.

[3] K. Pastra and Y. Wilks, "Image-language multimodal corpora: needs, lacunae and an ai synergy for annotation," in *Proceedings of the Language Resources and Evaluation Conference*, 2004, pp. 767–770.

[4] M. Ogiela, R. Tadeusiewicz, and L. Ogiela, "Graph image language techniques supporting radiological, hand image interpretations," *Computer Vision and Image Understanding*, vol. 103, no. 2, pp. 112–120, 2006.

[5] R. Mooney, "Learning to connect language and perception," in *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI)*, 2008, pp. 1598–1601.

[6] D. Pérez and S. Guadarrama, “Aprendizaje de la sintaxis para la descripción de escenas compuestas por figuras geométricas,” in *Proceedings ESTYLF 2010*, 2010, pp. 49–54.

[7] D. Chen and R. Mooney, “Learning to sportscast: A test of grounded language acquisition,” in *Proc. 25th International Conference on Machine Learning (ICML)*, 2008.

[8] S. Jaime-Castillo, J. Medina, and D. Sánchez, “Using FORDBMS for the linguistic description of images,” submitted to WCCI 2010.

[9] J. Chamorro-Martínez, J. Soto-Hidalgo, and D. Sánchez, “A new approach for defining a fuzzy color space,” submitted to WCCI 2010.

[10] A. Moghaddamzadeh and N. Bourbakis, “A fuzzy region growing approach for segmentation of color images,” *Pattern Recognition*, vol. 30, no. 6, pp. 867–881, 1997.

[11] J. Maeda, C. Ishikawa, S. Novianto, N. Tadehara, and Y. Suzuki, “Rough and accurate segmentation of natural color images using fuzzy region-growing algorithm,” in *15th International Conference on Pattern Recognition*, vol. 3, October 2000, pp. 638–641.

[12] S. Philipp-Foliguet, M. Bernardes Viera, and A. Albuquerque Araujo, “Segmentation into fuzzy regions using topographic distance,” *Proceedings of the XIV Brazilian Symposium on Computer Graphics and Image Processing*, pp. 282–288, 2001.

[13] P. Sobrevilla and E. Montseny, “Fuzzy sets in computer vision: An overview,” *Mathware & Soft Computing*, vol. 10, pp. 71–83, 2003.

[14] H. D. Cheng and J. Li, “Fuzzy homogeneity and scale-space approach to color image segmentation,” *Pattern Recognition*, vol. 36, no. 7, pp. 1545–1562, 2003.

[15] S. Makrogiannis, G. Economou, and S. Fotopoulos, “A region dissimilarity relation that combines feature-space and spatial information for color image segmentation,” *IEEE Transactions on Systems, Man & Cybernetics, Part B Cybernetics*, vol. 35 (1), pp. 44–53, 2005.

[16] B. Prados-Suarez, J. Chamorro-Martínez, D. Sánchez, and J. Abad, “Region-based fit of colour homogeneity measures for fuzzy image segmentation,” *Fuzzy Sets and Systems*, vol. 158, no. 3, pp. 215–229, 2007.

[17] Y. Haxhimusa and W. Kropatsch, “Hierarchical image partitioning with dual graph contraction,” in *Proc. of 25th DAGM Symposium LNCS*. Springer, 2003, pp. 338–345.

[18] J. C. Tilton, “Method for recursive hierarchical segmentation by region growing and spectral clustering with a natural convergence criterion,” 2000, disclosure of Invention and New Technology: NASA Case No. GSC 14,328-1.

[19] M. Jeon, M. Alexander, W. Pedrycz, and N. Pizzi, “Unsupervised hierarchical image segmentation with level set and additive operator splitting,” *Pattern Recognition Letters*, vol. 26, pp. 1461–1469, 2005.

[20] A. Kuijper and L. M. J. Florack, “The hierarchical structure of images,” *IEEE Transactions on Image Processing*, vol. 12, pp. 1067–1079, 2003.

[21] D. Saupe, M. Ruhl, R. Hamzaoui, L. Grandi, and D. Marini, “Optimal hierarchical partitions for fractal image compression,” in *IEEE Int. Conf. on Image Processing ICIP’98*, 1998.

[22] F. A. Taba, G. Naghdya, and A. Mertins, “Scalable multiresolution color image segmentation,” *Signal Processing*, vol. 86, pp. 1670–1687, 2006.

[23] Z. Tu and S. Zhu, “Parsing images into regions, curves and curve groups,” *International Journal of Computer Vision*, vol. 69, pp. 223–249, 2006.

[24] J. C. Tilton, G. Marchisio, and M. Datcu, “Knowledge discovery and data mining based on hierarchical segmentation of image data,” 2000, a research proposal submitted October 23, 2000 in response to NRA2-37143 from NASA’s Information Systems Program.

[25] S. Lhermitte, J. Verbesselt, I. Jonckheere, K. Nackaerts, J. A. van Aardt, W. W. Verstraeten, and P. Coppin, “Hierarchical image segmentation based on similarity of NDVI time series,” *Remote Sensing of Environment*, vol. 112, pp. 506–521, 2008.

[26] M. Liévin and F. Luthon, “Nonlinear color space and spatiotemporal mrf for hierarchical segmentation of face features in video,” *IEEE Transactions on Image Processing*, vol. 13, pp. 63–71, 2004.

[27] R. Hesami, A. BabHadiashar, and R. HosseinNezhad, “Range segmentation of large building exteriors: A hierarchical robust approach,” *Computer Vision and Image Understanding*, 2010, in press, doi:10.1016/j.cviu.2009.12.004.

[28] J. Chamorro-Martínez, D. Sánchez, B. Prados-Suárez, E. Galán-Perales, and M. Vila, “Segmenting colour images on the basis of a fuzzy hierarchical approach,” *Mathware & Soft Computing*, vol. 10, pp. 101–115, 2003.

[29] B. Prados-Suárez, D. Sánchez, and J. Chamorro-Martínez, “A similarity measure between fuzzy regions to obtain a hierarchy of fuzzy image segmentations,” in *Proceedings WCCI 2008*, 2008, pp. 1647–1654.

[30] A. Cohn and S. Hazarika, “Qualitative spatial representation and reasoning: An overview,” *Fundamenta Informaticae*, vol. 46, no. 1–2, pp. 1–29., 2001.

[31] N. E. Maillot and M. Thonnat, “Ontology based complex object recognition,” *Image and Vision Computing*, vol. 26, pp. 102–1131, 2008.

[32] U. Straccia, “Towards spatial reasoning in fuzzy description logics,” in *Proceedings Fuzz-IEEE 2009*, 2009, pp. 512–517.

[33] S. Schockaert, M. D. Cock, , and E. E. Kerre, “Spatial reasoning in a fuzzy region connection calculus,” *Artificial Intelligence*, vol. 173, no. 2, pp. 258–298, 2009.

[34] T. Regier, *The Human Semantic Potential*. Cambridge, MA: MIT Press, 1996.

[35] C. Hudelot, J. Atif, and I. Bloch, “Fuzzy spatial relation ontology for image interpretation,” *Fuzzy Sets and Systems*, vol. 159, pp. 1929 – 1951, 2008.

[36] J. Freeman, “The modelling of spatial relations,” *Comput. Graphics Image Processing*, vol. 4, no. 2, pp. 156–171, 1975.

[37] B. Kuipers and T. Levitt, “Navigation and mapping in large-scale space,” *AI Magazine*, vol. 9, no. 2, pp. 25–43, 1988.

[38] S. Dutta, “Approximate spatial reasoning: Integrating qualitative and quantitative constraints,” *International Journal of Approximate Reasoning*, vol. 5, pp. 307–331, 1991.

[39] K. Miyajima and A. Ralescu, “Spatial organization in 2d segmented images: Representation and recognition of primitive spatial relations,” *Fuzzy Sets and Systems*, vol. 65, pp. 23–38, 1994.

[40] R. K. Goyal and M. J. Egenhofer, “Similarity of cardinal directions,” in *Proceedings of the International Symposium on Advances in Spatial and Temporal Databases*. Springer-Verlag, 2001.

[41] I. J. Sledge and J. M. Keller, “Mapping natural language to imagery: Placing objects intelligently,” in *Proceedings Fuzz-IEEE 2009*, 2009, pp. 518–523.

[42] P. Matsakis and L. Wendling, “A new way to represent the relative position between areal objects,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 634–643, 1999.

[43] P. Matsakis, J. M. Keller, O. Sjahputera, and J. Marjamaa, “The use of force histograms for affine-invariant relative position description,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1–18, 2004.

[44] S. Winter, “Topological relations in hierarchical partitions,” in *Proceedings COSIT’99*, C. Freksa and D. Mark, Eds. Berlin-Heidelberg: Springer Verlag, 1999, pp. 141–155.

[45] M. Delgado, D. Sánchez, and M. Vila, “Fuzzy cardinality based evaluation of quantified sentences,” *International Journal of Approximate Reasoning*, vol. 23, pp. 23–66, 2000.

[46] K. Kelly and D. Judd, “The ISCC-NBS method of designating colors and a dictionary of color names,” *National Bureau of Standards (USA)*.

[47] ——, “Color universal color language and dictionary of names,” *National Bureau of Standards (USA)*, no. 440.

[48] B. Berlin and P. Kay, *Basic color terms: their Universality and Evolution*. Berkeley: University of California Press, 1969.

[49] R. Castillo-Ortega, N. Marín, and D. Sánchez, “Linguistic summary-based query answering on data cubes with time dimension,” in *FQAS’09*, ser. LNAI, T. A. et al., Ed., vol. 5822. Springer, Heidelberg, 2009, pp. 560–571.

[50] M. A. Vila, J. C. Cubero, J. M. Medina, and O. Pons, “The generalized selection: an alternative way for the quotient operations in fuzzy relational databases,” in *Fuzzy Logic and Soft Computing*, B. Bouchon-Meunier, R. Yager, and L. Zadeh, Eds. World Scientific Press, 1995.

[51] B. Neumann and R. Möller, “On scene interpretation with description logics,” *Image and Vision Computing*, vol. 26, pp. 82–101, 2008.