Towards a Similarity between Qualitative Image Descriptions for Comparing Real Scenes

Zoe Falomir *, Lledó Museros, Luis Gonzalez-Abril
Universität Bremen (Germany), Universitat Jaume I at Castelló (Spain), Universidad de Sevilla (Spain)

Abstract

In this paper, qualitative descriptors of shape, color, topology and location are applied to describe images of real scenes. Then, an approach for obtaining a similarity measure between two qualitative image descriptions (SimQID) is presented. As a proof-of-concept, this approach is tested on real digital images taken by a robot camera at the corridors of TI building at Universitat Jaume I. The results obtained are analyzed and a discussion is provided.

Introduction

Companion robots and ambient intelligent systems are designed to interact with human beings. Those systems usually integrate digital cameras from which they obtain information about the environment. The ideal systems for interacting with people would be those capable of interpreting their environment cognitively, that is, similarly to how people do it. In this way, those systems may align its concepts with human concepts and thus provide common ground for establishing sophisticated communication. However, although many advances have been carried out in computer vision, there is still research on scene understanding going on in order to answer questions such as: what has changed in the scene? or where is the difference in both images? What people can solve as a killer game, it is not so trivial for computer vision. However, qualitative spatial and temporal representations of the space can help towards formulating a cognitive solution.

In the literature, psychological studies explain that people find the relevant content in the images and use words (qualitative concepts) to describe it (Laine-Hernandez and Westman 2006; Greisdorf and O’Connor 2002; Wang et al. 2008). Usually different colored regions in an image indicate different objects of interest to people (Palmer 1999). Other studies explain also that, although the retinal image of a visual object may be considered a quantitative sensory input, the knowledge about the image that can be retrieved from memory is qualitative, since people report on what they saw using words and approximate terms (Freksa 1991a). Thus, qualitative representations of images are in many ways similar to the ‘mental images’ (Kosslyn, Thompson, and Ganis 2006) that people report when they attempt to answer questions on the basis of visual memories.

Because digital images represent visual data numerically, most image processing has been successfully carried out by applying mathematical techniques to obtain and describe image content. However, there is a growing trend of works in the literature which extract qualitative/semantic information from images. The following approaches described images semantically with only a word/concept: (i) using the shape of the Fourier transforms to extract perceptual properties of the images (i.e. naturalness, openness, etc.) for classifying them into semantic categories such as coast, mountain, etc. (Oliva and Torralba 2001); (ii) classifying images of indoor scenes in semantic categories (i.e. kitchen, bathroom, etc.) using a learning distance for object recognition and training on a dataset (Quattoni and Torralba 2009), etc. Other approaches described images semantically using more than one concept: (i) using a grid for dividing landscape images and assigning a semantic category (i.e. grass, water, etc.) to each cell (Qayyum and Cohn 2007); (ii) classifying images of natural scenes which may belong to more than one class (i.e. coast and mountain) using a fuzzy qualitative approach (Lim and Chan 2012); etc. All these approaches provide evidence for the effectiveness of using qualitative/semantic information to describe images. They all obtain a similarity index between images and use it to classify them into categories. However, very few approaches describe real images semantically as a set of components arranged in the space and use this description for classifying them. Some approaches found in the literature are applied to the retrieval of sketches or clip-art drawings (Sousa and Fonseca 2009; 2010), but not to real scenes, as far as we are concerned.

Furthermore, image similarity has also been used to solve the popular SLAM problem in robotics, obtaining successful results in worlds which contain textured objects that can be easily recognized by feature detectors (Ramisa et al. 2009). In environments containing low textured or incomplete objects, those object detectors do not perform so well and landmarks are more difficult to identify. However, qualitative descriptors can be used in those cases.
Qualitative Image Descriptions (QIDs) based on visual and spatial features (Falomir et al. 2012) showed high adaptability to different real-world scenarios and high flexibility for integration with: (i) description logics (Falomir et al. 2011) and (ii) qualitative distances (Falomir et al. 2013b). In this paper, a new measure of similarity for comparing two QIDs is presented. This measure extends previous works on qualitative shape and colour similarity (Falomir et al. 2013c; 2013a; 2013d) and the novel contributions of this paper are (1) the definition of a similarity measure between the spatial features of the objects in the images (topology and location) and (2) the formulation of a similarity measure between two images, which also obtains a correspondence of objects.

The rest of the paper is organized as follows. First, the approach for Qualitative Image Description (QID) is summarized. Then a measure of similarity for each of the described qualitative features is presented. Finally, a measure of similarity between QIDs is defined and tested and the results of the experimentation are shown and discussed.

**Qualitative Image Descriptors (QIDs)**

The QID approach (Falomir et al. 2012) applies a graph-based region segmentation method (Felzenszwalb and Huttenlocher 2004) to a digital image and then extracts the closed boundary of the relevant regions detected. Each object/region extracted is described qualitatively by describing its shape (QSD) and its color (QCD). To build the spatial representation, the object is considered to be positioned in the 2D image space, and its topological description and its location description are provided. Note that topology relations also implicitly describe the relative distance between the objects. Thus, the complete image is described by a set of qualitative descriptors of objects as:

\[[[\text{QSD}_1, \text{QCD}_1, \text{Topology}_1, \text{Orientation}_1], \ldots, [\text{QSD}_k, \text{QCD}_k, \text{Orientation}_k, \text{Topology}_k]]\]

where \(k\) is the total number of objects in an image.

**Qualitative Shape Description (QSD)**

The slope of the pixels within the object boundary are analyzed and the relevant points of shape are extracted. Each of these relevant points (\(\{P_0, P_1, \ldots, P_N\}\)) is defined by a set of four features \(<\text{EC}_P, \text{Ap}, \text{TC}_P, \text{L}_P, \text{C}_P>\), which are summarized below.

- The Edge Connection (EC) occurring at \(P\), described as: \(\{\text{line}, \text{line_curve}, \text{curve_line}, \text{curve_curve}, \text{curve_point}\}\);
- Angle (A) at point \(P\) (which is not a curvature_point) described by the qualitative tags: \(\{\text{very_acute, acute, right, obtuse, very_obtuse}\}\) defined on the intervals \(A_{INT} = \{(0, 40), (40, 85), (85, 95), (95, 140), (140, 180)\}\);
- Type of Curvature (TC) at the relevant point \(P\) (which is a curvature_point) described qualitatively by the tags: \(\{\text{very_acute, acute, semicircular, plane, very_plane}\}\) defined on the intervals \(TC_{INT} = \{(0, 40), (40, 85), (85, 95), (95, 140), (140, 180)\}\);
- Compared Length (L) of the two edges connected by \(P\), described qualitatively by: \(\{\text{much_shorter (ms), half_length (hl), a_bit_shorter (absh), similar_length (sl), a_bit_longer (abli), double_length (dl), much_longer (ml)}\}\) defined on the intervals \(L_{INT} = \{(0, 0.4), (0.4, 0.6), (0.6, 0.9), (0.9, 1.1), (1.1, 1.9), (1.9, 2.1), (2.1, \infty)\}\);
- Convexity (C) at the relevant point \(P\), described as: \(\{\text{convex, concave}\}\).

Thus, the complete shape of an object is described by a set of qualitative descriptions of relevant points as:

\[\left[\left[\text{EC}_1, \text{A}_1 | \text{TC}_1, \text{L}_1, \text{C}_1\right], \ldots, \left[\text{EC}_n, \text{A}_n | \text{TC}_n, \text{L}_n, \text{C}_n\right]\right] \]

where \(n\) is the total number of relevant points of the object.

**Qualitative Colour Description (QCD)**

The Red, Green and Blue (RGB) color channels are translated into Hue, Saturation and Lightness (HSL) coordinates, and from the HSL coordinates, a Qualitative Color Reference System is defined as: QCRS = \{UH, US, UL, QC\_LAB\_INT, QC\_INT\} where UH is the Unit of Hue; US is the Unit of Saturation; UL is the Unit of Lightness; QC\_LAB\_INT refers to the qualitative labels related to color; and QC\_INT refers to the intervals of HSL color coordinates associated with each color label. The chosen QC\_LAB\_INT and QC\_INT are:

- QC\_LAB\_INT = \{\text{black, dark-grey, grey, light-grey, white}\}
- QC\_INT = \{(0, 20), (20, 30), (30, 40), (40, 80), (80, 100) \in UL \lor UH \land US \in [0, 20]\}
- QC\_LAB = \{\text{red, orange, yellow, green, turquoise, blue, purple, pink}\}
- QC\_INT = \{(335, 360], [0, 15], (15, 40], (40, 80], (80, 160], (160, 200], (200, 260], (260, 297], (297, 335) \in UH \lor US \in \{50, 100\} \land UL \in (40, 55]\}
- QC\_LAB = \{\text{pale-} + QC\_LAB\}
- QC\_INT = \{\forall UH \land U \in (20, 50] \land UL \in (40, 55]\}
- QC\_LAB = \{\text{light-} + QC\_LAB\}
- QC\_INT = \{\forall UH \land U \in (50, 100] \land UL \in (55, 100]\}
- QC\_LAB = \{\text{dark-} + QC\_LAB\}
- QC\_INT = \{\forall UH \land U \in (20, 40]\}

The QCRS depends on the vision system used. However, it can be adapted to other systems and/or scenarios by defining other color tags and/or other HSL values (Falomir et al. 2013c).

**Topological Description**

In order to represent the topological relations of the objects in the image, a model defined for region configurations in \(R^2\) is used (Egenhofer and Franzosa 1991), which describes the topology situation in space (invariant under translation, rotation and scaling) of an object A with respect to (wrt) another object B (A wrt B) as:

\[T_{LAB} = \{\text{disjoint, touching, completely_inside, container}\}\]

The \(T_{LAB}\) determines if an object is completely_inside or if it is the container of another object. It also defines the neighbors of an object as all the other objects with the same

\(\text{Arc}_i\) denotes the angle or the type of curvature that occurs at the point \(P_i\).
A Measure of Similarity between QIDs

In order to define a similarity measure between two Qualitative Image Descriptions (QIDs), it is necessary to define a similarity measure between:

1. Qualitative Shape Descriptions ($SimQSD$),
2. Qualitative Color Descriptions ($SimQCD$),
3. Topology Descriptions ($SimTop$),
4. Location Descriptions ($SimLo$),
5. all the objects contained in both compared images which combines the previous defined similarities ($SimQID$).

Similarity of Shape ($SimQSD$)

For obtaining a similarity measure between QSDs, three previous definitions are needed: (1) Similarity between qualitative tags related to the same feature of shape; (2) Similarity between relevant points; and (3) Similarity between QSDs by correspondence of their relevant points (Ps).

Similarity between Qualitative Features

This approach uses two measures of similarity between qualitative terms: (i) based on conceptual neighborhood diagrams, for pure qualitative terms: EC and C; and (ii) based on interval distances for the qualitative terms defined as a discretization of a value on a reference system: A, TC and L.

Two qualitative terms are conceptual neighbors if ‘one can be directly transformed into another by continuous deformation’ (Freksa 1991b). For example, any connection involving two curves can evolve directly to a curvature-point but not to a line-line segment, because first it has to evolve to a curve-line or line-curve connection.

In general, Conceptual Neighborhood Diagrams or CNDs are described as diagrams or graphs containing: (i) nodes that map to a set of individual relations defined on regions
or intervals and (ii) paths or edges connecting pairs of adjacent nodes that map to continuous transformations between them. Figure 3 shows the CND proposed for the feature EC and Table 1 shows the dissimilarity between each label calculated from the CND as the minimal path between them. The same is done for the feature C, whose CND is trivial.

![CND for feature Kind of Edges Connected](image)

Table 1: Dissimilarity matrix for EC using a CND.

<table>
<thead>
<tr>
<th>EC</th>
<th>line</th>
<th>line</th>
<th>curve</th>
<th>curve</th>
<th>curve</th>
<th>curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>line</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>curve</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>curve</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>curve</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>curvature</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

The qualitative names for the features of shape A, TC and L are defined from intervals of values in their Reference Systems. Hence, interval distances can be used to measure the dissimilarity between them.

Let us introduce the concept of interval distance. Given an open interval (analogously for another kind of interval) of finite dimension, there are two main ways to represent it: from the extreme points as $(a,b)$ or as an open ball $B_r(c)$ (Borelian notation) where $c = (a + b)/2$ (centre) and $r = (b - a)/2$ (radius). Given two intervals, $I_1 = (a_1, b_1)$ and $I_2 = (a_2, b_2)$, a family of distances between intervals is defined (Gonzalez-Abril et al. 2004) as follows:

$$d^2(I_1, I_2) = \left( \frac{\Delta c}{\Delta r} \right) A \left( \begin{array}{c} \Delta c \\ \Delta r \end{array} \right)$$

where $\Delta c = c_2 - c_1$, $\Delta r = r_2 - r_1$ and $A$ is a symmetrical $2 \times 2$ matrix of weights that must be a positive definite matrix. From the A matrix, the weights given to the position of the intervals and to the radius can be controlled. In this paper, the identity matrix is used that provides the next distance:

$$d(I_1, I_2) = \sqrt{\Delta c^2 + \Delta r^2} = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2}$$

(1)

Hence, applying (1) to $A_{INT}$ and $TC_{INT}$ and $L_{INT}$, matrices of dissimilarities between the qualitative names in $A_{LAB}$, $TC_{LAB}$ and $L_{LAB}$ can be obtained. Note that the intervals in $A_{INT}$ and $TC_{INT}$ have different amplitudes and that interval distances take into account the amplitude of the compared intervals to calculate the dissimilarity between two qualitative names.

**Similarity between Relevant Points (Ps)** As the qualitative shape of an object is described by its relevant points, a similarity between them is needed. Hence, given two relevant points, denoted by $P_A$ and $P_B$, belonging to the shapes of the objects A and B respectively, a similarity between them, denoted by $Sim(P_A, P_B)$, is defined as:

$$Sim(P_A, P_B) = 1 - \sum_{i \in \{EC, A, V, TC, C, L\}} w_i \frac{ds(i)}{Ds(i)},$$

(2)

where $ds(feature)$ and $Ds(feature)$ denote the dissimilarity between relevant points and the maximum dissimilarity respectively wrt the feature obtained from the dissimilarity matrix. Hence, by dividing $ds(feature)$ and $Ds(feature)$ the proportion of dissimilarity related to feature of $P_A$ and $P_B$ is obtained, which is between 0 and 1. Moreover, the weight $w_{feature}$ is the weight assigned to this feature, and it is hold that $w_{EC} + w_A + w_L + w_C = 1$, $w_A = w_{TC}$ and $w_{feature} \geq 0$ for each feature.

In this paper, with the aim of giving the same importance to all features in (2), all the weights have the same value: $\frac{1}{4}$. Clearly, these weights can be tuned if a researcher needs to give more importance to one feature over the others. Furthermore, in (2) it is subtracted from 1 with the aim of providing a similarity instead of a dissimilarity.

For each $P_A$ and $P_B$, it is straightforward to prove that $0 \leq Sim(P_A, P_B) \leq 1$ and that this is a symmetrical relation. Furthermore, $Sim(P_A, P_B) = 0$ means that both relevant points are as different as possible, and $Sim(P_A, P_B) = 1$ means that both relevant points are the same.

**Similarity of QSDs by Correspondence of Ps** In order to compare two shapes A and B whose QSDs have the same number of relevant points (denoted by $m$), the similarity between A and B, denoted by $SimQSD(A, B)$, is calculated from (2) as follows: Fixing a relevant point of $A$, $P_A$, $i = 1, \cdots, m$, the similarities between the pairs of relevant points of the set

$$C_i = \{(P_A^1, P_B^1), \cdots, (P_A^m, P_B^m+1-i),$$

$$\cdots, (P_A^i, P_B^{m+2-i}), \cdots, (P_A^{m-1}, P_B^m)\}$$

are calculated. Thus,

$$SimQSD(A, B) = \max_{i=1, \cdots, m} \left\{ \frac{1}{m} \sum_{(P_A^i, P_B^i) \in C_i} Sim(P_A, P_B) \right\}$$

Let us see an example for two triangles, $T_A$ and $T_B$. If $T_A = \{P_A^1, P_A^2, P_A^3\}$ and $T_B = \{P_B^1, P_B^2, P_B^3\}$, then three similarities are considered (for simplifying, we denote $Sim(P_A^i, P_B^i)$ as $Sim(i, j)$):

$$Sim_1(T_A, T_B) = \frac{1}{3} (Sim(1, 1) + Sim(2, 2) + Sim(3, 3))$$

$$Sim_2(T_A, T_B) = \frac{1}{3} (Sim(2, 1) + Sim(3, 2) + Sim(1, 3))$$

$$Sim_3(T_A, T_B) = \frac{1}{3} (Sim(3, 1) + Sim(1, 2) + Sim(2, 3))$$
and, the final similarity between \( T_A \) and \( T_B \) is the maximum of these three. Note that a correspondence between relevant points of two shapes is provided with this similarity: if the final similarity between the triangles \( T_A \) and \( T_B \) is given from the \( \text{Sim}_2(T_A, T_B) \), then the correspondence obtained is: \( P_A^1 \rightarrow P_B^1, P_A^2 \rightarrow P_B^2, \) and \( P_A^3 \rightarrow P_B^3. \)

In general, if the number of relevant points of the shapes \( A \) and \( B \) are \( n \) and \( m \) respectively (and assuming without loss of generality that \( n \geq m \)), then there are \( n - m \) relevant points of \( A \) shape with no corresponding points in the \( B \) shape.

Let consider \( C \) the set of all possible ways (combinations) to chose \( n - m \) relevant points of \( A \). Hence, if \( c \in C \), a new shape \( A' \) is considered such that \( A' \) is given by all the relevant points of \( A \) without the \( n - m \) points without correspondence in \( B \). Hence \( A' \) and \( B \) have the same number of relevant points and its similarity can be calculated as in the previous case.

Thus, the similarity between \( A \) and \( B \) is obtained as:

\[
\text{Sim}_{QSD}(A, B) = \frac{m}{n} \max_{c \in C} \{ \text{Sim}_{QSD}(A', B) \}
\]  

**Similarity of Colour \((\text{Sim}_{QCD})\)**

The qualitative color names \((QC)\) are defined in HSL from three intervals \( I^H \times I^S \times I^L \), as follows: \( QC = \{h_0, h_1\} \times [s_0, s_1] \times [l_0, l_1] = B_r(x, y) \times B_r(lc) \)

Hence, given two color names \( QC_A = B_r(x_1, y_1) \times B_r(lc_1) \) and \( QC_B = B_r(x_2, y_2) \times B_r(lc_2) \), the interval distance between them is calculated from (1) as:

\[
d_{QCD}(QC_A, QC_B) = d_{(1,0.5)}(B_r(x_1), B_r(x_2)) + d_{(1,0.5)}(B_r(y_1), B_r(y_2)) + d_{(2,2)}(B_r(lc_1), B_r(lc_2))
\]

Clearly, \( d_{QCD} \) is a distance since it is defined from an addition of distances. Hence, given two color names \( QC_A \) and \( QC_B \), a normalized similarity \( \text{Sim}_{QCD} \) between them is defined as:

\[
\text{Sim}_{QCD}(QC_A, QC_B) = 1 - \frac{d_{QCD}(QC_A, QC_B)}{\max(d_{QCD})}
\]  

where \( \max(d_{QCD}) \) denotes the maximum distance for all colors defined by \( QC \). Hence, by dividing \( d_{QCD}(QC_A, QC_B) \) and \( \max(d_{QCD}) \) the proportion of dissimilarity related to qualitative colors \( QC_A \) and \( QC_B \) is obtained, which has values between 0 and 1. Finally, this value is subtracted from 1 with the aim of providing a similarity instead of a dissimilarity (Falomir et al. 2013c).

**Similarity of Topology \((\text{Sim}_{Top})\)**

When describing the topological situation of two objects, \( A \) and \( B \), located respectively in two different images, \( Im_A \) and \( Im_B \), the QID approach describes is:

- the **containers** that objects \( A \) and \( B \) have,
- the components that \( A \) and \( B \) have (the other objects located **completely inside** \( A \) and \( B \)),
- the **other objects** which are **touching** \( A \) and \( B \), and
- the other objects which are **disjoint** \( A \) and \( B \).

In order to calculate if the topological situation of objects \( A \) and \( B \) is similar or not, this approach takes into account the four features described: (1) quantity (\( Q \)) of containers, (2) quantity of objects **completely inside**, (3) quantity of neighbors **touching** and (4) quantity of neighbors **disjoint** which each object has, rather than which objects are exactly.

In this way, the degree of similarity obtained depends only in the topological situation in the space of a specific object (\( A \)) wrt another specific object (\( B \)).

Therefore, the similarity of two objects \( A \) and \( B \) wrt a topological relation \( x \), is defined as:

\[
\text{Top}(x, A, B) = \begin{cases} 
\frac{1}{\text{Max}_{QSD}(x, A, B)} & \text{if } Q(x, A) = Q(x, B) \\
\text{otherwise} & \\
\end{cases}
\]

where \( x = \{ \text{Container, Completely Inside, Neighbor Disjoint, Neighbor Touching} \} \), \( Q \) is the quantity of \( x \), \( \text{Min}_Q \) is the minimum quantity of \( x \) of \( A \) and \( B \), and finally \( \text{Max}_Q \) is the maximum quantity of \( x \) of \( A \) and \( B \). Note that it is hold that \( 0 \leq \text{Top}(x, A, B) \leq 1 \) because \( \text{Min}_Q(x, A, B) \leq \text{Max}_Q(x, A, B) \).

Therefore, the similarity of two objects \( A \) and \( B \), wrt these topological features is:

\[
\text{Sim}_{Top}(A, B) = w_C \cdot \text{Top}(\text{Containers}, A, B) + w_{CI} \cdot \text{Top}(\text{Completely Inside}, A, B) + w_D \cdot \text{Top}(\text{Neighbors Disjoint}, A, B) + w_T \cdot \text{Top}(\text{Neighbors Touching}, A, B)
\]  

where \( w_C + w_{CI} + w_D + w_T = 1 \) and, therefore, it is straightforward to prove that \( 0 \leq \text{Sim}_{Top}(A, B) \leq 1 \).

**Similarity of Location \((\text{Sim}_{Lo})\)**

From the LoRS in Figure 2, the CND showed in Figure 4 can be built. The weights in this CND have been calculated in order to get the maximum dissimilarity between opposite location relations (i.e. up vs. down, right vs. left, up-left vs. down-right etc.). Therefore, the dissimilarity matrix obtained from this CND is that shown in Table 2. Thus, given two locations, denoted by \( Lo_A \) and \( Lo_B \), referring to the locations of the objects \( A \) and \( B \) respectively, a similarity between them, denoted by \( \text{Sim}_{LoRel}(Lo_A, Lo_B) \), is defined as:

\[
\text{Sim}_{LoRel}(Lo_A, Lo_B) = 1 - \frac{d_{\text{LoRel}(Lo_A, Lo_B)}}{\text{Max}_{DSimLo}}
\]  

Figure 4: CND for the LoRS.
where \( dsLoRel(Lo_A, Lo_B) \) denotes the dissimilarity between the locations obtained from the dissimilarity matrix previously defined. \( MaxDSimLo \) denotes the maximum dissimilarity for all locations (i.e. 4 for this case of study). Hence, by dividing \( dsLoRel(Lo_A, Lo_B) \) and \( MaxDSimLo \) the proportion of dissimilarity related to locations \( Lo_A \) and \( Lo_B \) is obtained, which has values between 0 and 1. This value is subtracted from 1 with the aim of providing a similarity.

If the quantity of locations of the objects A and B are \( n \) and \( m \) respectively, and \( n \geq m \). In this case, \( n-m \) locations of A are compared with the \( void \) orientation and they similarity is 0, and the rest are compared with the locations of B as shown in (6). Therefore, the similarity of two objects A and B, with respect to their locations is:

\[
SimLo(A, B) = \frac{\sum_{Lo_A \in A, Lo_B \in B} SimLoRel(Lo_A, Lo_B)}{n} \tag{7}
\]

**Similarity of Images (SimQID)**

In order to define a similarity measure between images, first a similarity between objects must be obtained. Hence, given two objects, denoted by A and B, a similarity between them, denoted by \( SimObj(A, B) \), is defined as:

\[
SimObj(A, B) = w_{QSD} \cdot SimQSD(A, B) + w_{QCD} \cdot SimQCD(A, B) + w_{Top} \cdot SimTop(A, B) + w_{Lo} \cdot SimLo(A, B) \tag{8}
\]

where the parameters \( w_{QSD}, w_{QCD}, w_{Top} \) and \( w_{Lo} \) are the weights assigned to the shape similarity (\( SimQSD \)), the color similarity (\( SimQCD \)), the topology similarity (\( SimTop \)) and the location similarity (\( SimLo \)), respectively. Moreover, it is hold that \( w_{QSD} + w_{QCD} + w_{Top} + w_{Lo} = 1 \). Clearly, these weights can be tuned in order to give more importance to one feature (shape, color, topology or orientation) over the others. Finally, for each A and B, it is straightforward to prove that \( 0 \leq SimObj(A, B) \leq 1 \) and that this is a symmetrical relation.

Therefore, in order to compare two images \( Im_A \) and \( Im_B \) whose QIDs have the same number of objects (denoted by \( N \)), the similarity between \( Im_A \) and \( Im_B \) is calculated from (8) as an arithmetic mean of the similarity between objects:

\[
SimQID(Im_A, Im_B) = \frac{1}{n} \sum_{A \in Im_A} SimObj(A, B) \tag{9}
\]

Note that, depending on which objects of \( Im_A \) are compared with which objects of \( Im_B \), different values of \( SimQID \) are obtained. The final correspondences between objects obtained by our approach will be those which maximize the similarity between images, because this will mean that each object has been matched to the most similar one:

\[
SimQIDFinal(Im_A, Im_B) = \max_C(SimQID(Im_A, Im_B)) \tag{10}
\]

The Branch and Bound technique was used for speeding the calculus of (10). If the two images compared, \( Im_A \) and \( Im_B \), have a different number of objects, then there are some objects of one image with no corresponding objects in the other image. In this case, the objects with no corresponding pairs in the other image are compared with the void object, and the similarity between both objects is zero.

Let us suppose that the number of objects of the images \( Im_A \) and \( Im_B \) are \( N \) and \( M \) respectively, and that \( N \geq M \). In this case, \( N-M \) objects of \( Im_A \) are compared with the void object, and the rest are compared with the objects of \( Im_B \) in the same way as in the previous case. Taking into account this situation, the similarity between images is obtained from (9) and (10) in the same way as before.

### Experimentation and Results

As a proof-of-concept for the SimQID approach, let us consider the following scenario. A Pioneer robot navigates through the corridors of the TI building at Universitat Jaume I (Figure 5) and takes the photos showed in Table 3. This table also presents the objects described in each scene.

![Figure 5: Pioneer robot at UJI corridors](image)

<table>
<thead>
<tr>
<th>Location</th>
<th>u</th>
<th>ur</th>
<th>r</th>
<th>dr</th>
<th>d</th>
<th>dl</th>
<th>l</th>
<th>ul</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>up (u)</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>up-right (ur)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>right (r)</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>down-right (dr)</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>down (d)</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>down-left (dl)</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>left (l)</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>up-left (ul)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>center (c)</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 3: Comparing images of obtained from the robot camera.

![Image 1](img1) ![Image 2](img2) ![Image 3](img3) ![Image 4](img4)

**Table 4: Results of SimQID values and correspondences of objects obtained from the images in Table 3 with the \( w_{QSD} = w_{QCD} = 0.35, w_{Top} = w_{Lo} = 0.15 \).**

<table>
<thead>
<tr>
<th>Image</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.98</td>
<td>Image 1:{0, 1, 2, 3, 4, 5}</td>
<td>Image 1:{0, 1, 2, 3, 4, 5, ∅}</td>
</tr>
<tr>
<td></td>
<td>Image 2:{0, 1, 2, 3, 4, 5}</td>
<td>Image 3:{0, 1, 2, 3, 5, 6, 4}</td>
<td>Image 4:{0, 1, 2, 3, 5, 6, 2}</td>
</tr>
<tr>
<td>2</td>
<td>0.80</td>
<td>Image 2:{0, 1, 2, 3, 4, 5, ∅}</td>
<td>Image 3:{0, 1, 2, 3, 4, 5, 0}</td>
</tr>
<tr>
<td></td>
<td>Image 3:{0, 1, 2, 3, 5, 6, 4}</td>
<td>Image 4:{0, 1, 4, 5, 3, 6, 2}</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.89</td>
<td>Image 3:{0, 1, 2, 3, 4, 5, 6}</td>
<td>Image 4:{0, 1, 2, 3, 4, 5, 6}</td>
</tr>
</tbody>
</table>

- image 2 and image 3 are 0.80 similar. The correspondences of the objects that change are: (i) object 4 in image 2 is object 5 in image 3 and (ii) object 4 in image 3 does not correspond to any object in image 2, so it is an extra/new object. When comparing image 1 and image 3, the similarity is 0.78. The correspondences of objects are the same as before, but the change in topology and location in object 3 decrease the similarity.

- image 3 and image 4 are 0.89 similar. The correspondences of the objects in the images are the same. However, object 5 in image 3 has disappeared and object 5 in image 4 has appeared as a new object. The correspondence of these two objects which have differences in color, topology and location, decreases the similarity. But still, the similarity between image 3 and 4 is higher than the similarity between image 4 and images 1 and 2, because image 4 has an object more and the location and topology of them are more different. Note that when comparing image 4 to image 2, the extra object in image 4 is object 4. But note that when comparing image 4 to image 1, the extra object or the object that changed in image 4 is object 2 since object 4 has been mapped to object 2 in image 1 because of their location.

In the current experimentation, the threshold could be decreased until 0.80 and similar images were obtained.

**Discussion and Future Work**

The presented approach calculates a similarity measure between two images described qualitatively and it also provides a correspondence between the objects in the images. This correspondence between the objects may be used for identifying the changes between the images and those changes may be explained using the qualitative description on shape, color, topology or location of the objects. These explanations are intended as future work.

Note that this approach does not need training. Moreover, it can be applied to images with low textured objects or to images containing not complete objects (i.e. parts of doors, floor, etc.)

More testing using real images of the same place from different point of views is required to study if the SimQID can help to solve the SLAM problem in low textured worlds. Further experimentation is also need to determine the minimum value required for the threshold to retrieve similar images from data bases.

As future work, we intend to define a similarity measure between scenes in the 3D physical space captured by a MS
Kinect. The expected performance of the approach would be the same or improved, since more spatial situations between objects will be differentiated (i.e. occluding in 3D vs. touching in 2D).

Finally, a psychological test is also intended as future work to study how the weights \( w_i \) can be tuned so that the SimQID approximates human perception.

**Acknowledgments**

This work was supported by European Commission through FP7 Marie Curie IF actions under project COGNITIVE-AM (GA 328763), the interdisciplinary Transregional Collaborative Research Center Spatial Cognition SFB/TR 8, Andalusian Regional Ministry of Economy (project SIMON Tlc-8052), Spanish Ministry of Economy and Competitiveness (project TIN2011-24147), Generalitat Valenciana (project GVA/2013/135) and Universitat Jaume I (project P11B2013-29).

**References**


