Semantic Feature Extraction Using Genetic Programming in Image Retrieval

Qingyong Li          Hong Hu            Zhongzhi Shi

Institute of Computing Technology-Chinese Academy of Sciences – Beijing, China
{liqy, huhong, shizz}@ics.ict.ac.cn

Abstract

One of the big hurdles facing current content-based image retrieval (CBIR) is the semantic gap between the low-level visual features and the high-level semantic features. We proposed an approach to describe and extract the global texture semantic features. According to Tamura texture model, we utilize the linguistic variable to describe the texture semantics, so it becomes possible to depict the image in linguistic expression such as coarse, fine. And we use genetic programming to simulate the human visual perception and extract the semantic features value. Our experiments show that the semantic features have good accordance with the human perception, and also have good retrieval performance. In some extent, our approach bridges the semantic gap in CBIR.

1. Introduction

In content-based image retrieval (CBIR) fields, the techniques about low-level features such as color, texture, and shape have been extensively researched, but their performance is not satisfactory because of the semantic gap between the low-level features and high-level concepts (textual description). In order to conquer this question, image retrieval systems should be aimed to support high-level (semantics-based) querying and browsing. Obviously, the description and extraction of image semantic is necessary step to semantic-based image retrieval.

There are two traditional methods to extract the image semantic feature. Firstly, semantic feature can be extracted based on image processing and domain knowledge [4]. There are three key processes: image segmentation, object recognition, and object relation analysis, but image segmentation and object recognition are a difficult problem. Obviously, there are holdbacks for this way; Secondly, we can get image semantics from external, such as manual labels, human interaction. But it is labour-intensive and subjective.

Psychological research shows that the global feature is always prior to the local feature in human visual perception [5]. Based on these psychological results, we focus on the global semantic feature, especially for the texture semantic features.

In this paper, we propose a global texture semantic described by linguistic variable [2] according to Tamura texture model [1]. Then we use genetic programming [3] to simulate the human visual perception and to extract the image texture semantics based on the image visual texture features. The diagram is illustrated in Figure 1. The description and extraction methods have two salient advantages:

(1) We can use linguistic expression to depict the image content.
(2) Genetic programming can successfully construct the map from the low-level feature to the high-level semantic feature, though without the domain knowledge.

The rest of the paper is organized as follows: the following section introduces the semantic feature description using linguistic variable; the next section illustrates the semantic extraction method by genetic programming. And the experiment results are showed in section 4. At last we make conclusion.

![Figure 1. The diagram of our semantic feature extraction method](image)

2 Semantic feature description using linguistic variable

Linguistic variable is a variable whose values are natural language expressions referring to some quantity of interest. These natural language expressions name for fuzzy sets composed of the possible numerical values that the quantity of interest can be assumed. Linguistic variable can be defined as below.

A linguistic variable is characterized by a quintuple denoted by \(< x, T(x), U, G, M >\). Here, \(x\) is the name of the variable, and \(T(x)\) is the set of the linguistic values of...
variable represents the texture's coarseness feature, using linguistic variable with example following we'll show how to describe the texture semantic describes (a) the set of natural language expressions that variable consists of two parts: (1) a dimension linguistic vector directionality to be our semantic features. Actually, the former three features are the most frequently used, so we select the coarseness, contrast and directionality to be our semantic features.

Texture semantic features are described by a 3-dimension linguistic vector \( X = (X_1, X_2, X_3) \), \( X_1 \) represents the texture’s coarseness feature, \( X_2 \) represents the contrast, and \( X_3 \) represents the directionality. In the following we’ll show how to describe the texture semantic using linguistic variable with example \( X_1 \).

The coarseness describes the granularity of the texture model. We always depict the texture in such words as very coarse, coarse, fine. Now we utilize the linguistic variable \( X_1 = \langle X_1, T(X_1), U, G, M \rangle \) where

- \( X_1 \), assumed as ‘coarseness’, is the name of the linguistic variable.
- \( T(X_1) \) is the set of the natural language terms which may be designated to \( X_1 \) as values. \( T(X_1) \) comprises the basic term set and extended term set. We designate the basic term set with word set \{‘coarse’, ‘fine’\}, and through the syntax rule \( G \) we’ll generate the extend term set, such as \{‘very coarse’, ‘somewhat fine’\}.
- \( U \) is the universe of discourse. Here we use the texture spectrum feature vector proposed by Wan [6] to approximate the image. It is a 256-dimension low-level visual feature by analyzing the intensity changes of the neighboring pixels in a 3*3 neighborhood. So the 256-dimension Euclidean space becomes the universe of discourse.
- \( G \) is a syntax rule which generates the extend term set. For example, through basic term ‘coarse’ and modifiers-defined in \( G \) named very, somewhat, not, etc... we can get extend terms: very coarse, somewhat coarse, not coarse etc...
- \( M \) is the semantic rule mapping from \( T \) to the set of fuzzy subsets of \( U \), that is to say, for \( u \in U \), \( t \in T(X_1) \), assuming a value in \([0, 1]\) called degree of membership to \( u \) and \( t \).

Examined from another viewpoint, a linguistic variable consists of two parts: (1) a syntactic part which describes (a) the set of natural language expressions that are the values of the linguistic variable as well as (b) the structure of set, and (2) a semantic part that associates each natural language expression with a fuzzy subset. The syntactic parts of the above formal definition are \( T \) and \( G \); the semantic parts are \( M \) and \( U \).

3. Semantic feature extraction by GP

It is important for the linguistic variable how to construct the semantic rule. \( M \) assigns the elements of \( T \) with a fuzzy set which measures the extent to which the texture image represented by a texture spectrum vector is in accordance with the linguistic value. In other viewpoint, it can be regarded as image comprehension – simple and basic comprehension. Thus, how to simulate the function of the simple visual comprehension becomes the task of the semantic rule \( M \).

As we well know, Genetic Programming (GP)[3] has potential as modeling and optimization techniques for complex real-word problems. GP, based on the powerful principle of “survival of the fittest”, model some natural phenomena of genetic inheritance and Darwinian strife for survival. The GP is also an example of a weak method, which makes few assumptions about problem domain. In this subsection, we present our GP-based algorithm for constructing the semantic rule \( M \).

Firstly, we formally describe the task as the following: through sample set \{(\( V_1, y_1 \), … , \( V_n, y_n \) \} in which \( V_i \) (\( i=1,...,n \)) is visual feature- 256 dimension texture spectrum vector, and \( y_i \) (\( i=1,...,n \)) is the degree of membership in \([0, 1]\) regarding to a special linguistic value, we learn to construct a function (or computable program):

\[
M_1 : V \rightarrow y \quad t \text{ is the linguistic value.}
\]

We select 200 images from Brodatz [7] texture database as the training sample set, label the degree of membership by five people according to special linguistic value, and designate the average as the final degree of membership.

Next, we shall address the most important three issues about GP: encoding scheme, fitness function and genetic operations.

- **Encoding scheme.** GP utilizes tree structures to represent the function as figure 2. The intra-node is always constituted by the basic functions such as plus, minus; and the leaf-node exclusively includes the terminate symbols, which is comprised of variable expression (composed of variable, coefficient, and power, for example \( x, x^3 \)) or constant. Obviously, each tree corresponds to a computable function.
Figure 2. Tree structure of polynomial $2X_1-5X_2+12X_3^2+0.5X_4$

- Fitness function. Fitness function influences the quality of the result. We define the fitness function as the root of the square sum of the error between the labeled degree of membership and evolution function result. Suppose that $F_i(v)$ represents the i-th individual, and the j-th training sample’s degree of membership is $\delta_j$, and the number of sample is n, then, we can get the fitness function:

$$Fit(F(v)) = \sqrt{\sum_{i=1}^{n} (F(v_j) - \delta_i)^2}$$

- Genetic operations. The tree structure allows easily define a closed crossover operator by swapping sub-trees between two valid trees. The scheme is the steady state genetic algorithm: parents are selected by tournament (of size of 2 to 7 typically), and an offspring is generated by crossover. The offspring is then put back in the population using a death-tournament: the individual with the worse fitness gets replaced by the new born offspring. We also use mutation operation: random replacement of a subtree or random change of a node or a leaf. The mutation probability is 0.02 in our experiment.

Generally speaking, each linguistic value of the linguistic variable has a corresponding semantic rule. But we can get the extend term’s semantic rule M through corresponding syntax rule G since the extend term is generated by the basic term and syntax rule G.

4. Experiment result

In this section, we’ll demonstrate the experiment result about the performance of the image semantic feature and the influence of the fitness in GP to retrieval performance. The images in this paper are selected from Brodatz[7] database.

4.1. Texture semantic feature and retrieval

In this paper, texture semantic features are described by a 3-dimension linguistic vector $X = (X_1, X_2, X_3)$. $X_1$ represents the texture’s coarseness feature, it’s basic term set is \{coarse, fine\}; $X_2$ represents the contrast, it’s basic term set is \{high, low\}; and $X_3$ represents the directionality, it’s basic term set is \{linear, irregular\}. Considering the storing cost, we only keep the basic term’s fuzzy values in the feature database, and the extended term’s values will be calculated through the syntax rule G and operators. So every image semantic feature comprises of a 6-dimension vector $SF=<v_1, v_2, v_3, v_4, v_5, v_6>$, the corresponding semantic is <coarse, fine, high, low, linear, irregular>.

![Image semantic feature description](image)

The images in figure 3 are randomly selected from our feature database. From figure 3, we can see that: for each semantic term, if the corresponding value or degree of membership is bigger, the image is more accordant with the semantic in perception. For example in figure 3, the left image’s fifth complement is 0.94, and that of the right is 0.07, in the same time we can feel that the left one is greatly more linear than the right one.

After we extract the texture semantic feature, we can retrieval images accord to the semantic feature in natural language. Follow figure illustrates the semantic retrieval interface and the retrieval result of the request “coarse and somewhat high contrast or somewhat linear”.

![Retrieval result of “coarse and somewhat high contrast or somewhat linear”](image)
‘high contrast’ and their logic combination. Obviously, it is more accustomed for the human, especially for browsing the image database to find an appropriate example. Furthermore, the semantic feature is tidy with low dimension.

4.2. The relationship between retrieval performance and GP

In our semantic-based image retrieval system, the semantic rule mapping the texture spectrum feature to the global texture semantic feature directly influences the retrieval performance measured by precision here. In GP, the quality of the semantic rule is measured by the fitness. When we construct the semantic rule using GP we assume that less the individual’s fitness, better its quality. But in effect, the relation between the individual’s fitness and retrieval performance is showed in Figure 5 with x axis being the GP result’s fitness and y axis being the system’s retrieval precision.

![Figure 5: The relation between the GP individual fitness and retrieval precision](image)

Figure 5 shows that the retrieval precision increases as the fitness decreases in the case of fitness greater than 0.5; But when it is less than 0.5, the precision doesn’t always ascend, on the contrary, it decreases in the point x=0.1. It is because that the fitness represents the training error in GP. In some extent, less the fitness is, better the GP is trained, so we can achieve the high retrieval precision. But the visual texture feature is an approximate of the image, and the training set and label semantic value used in GP are also subjective. Obviously, there are some limitations in precisely representing an image. So when the fitness is less than some value, the retrieval precision can’t achieve absolute increase as the fitness decreases. In our experiments, when the fitness equals 0.5 we can get the best retrieval precision.

5. Conclusion

In this paper, we propose a global image texture semantic feature described by linguistic variable according to Tamura texture model. Especially, we successfully utilize the genetic programming to simulate the human visual perception and extract the global image semantic feature. The experiment results show that the global image semantic feature can effectively capture the image's texture semantic and has a good accordance with the human perception. As for a global semantic feature, according to Navon's[5] "forest before tree" strategy, it is effective and efficient in image retrieval and browsing.

Reference