International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 11 | Nov 2019 www.irjet.net p-ISSN: 2395-0072

SAR IMAGE RETRIEVAL BASED ON REGION BASED SIMILARITY MEASURE FOR EARTH OBSERVATIONS

RACHIT JAIN[1], SUMIT YADAV[2], IRFAN MINSURWALE[3], SANDHYA SHINDE[4]

[1,2,3]Student, Dept. of Electronics And Tele-Communication Engineering, Dr. D.Y Patil Institute Of Engineering, Management and Research, Maharashtra, India.
[4]Assistant Professor, Dept. of Electronics and Tele-Communication Engineering, Dr. D.Y Patil Institute of Engineering, Management and Research, Maharashtra, India.

Abstract - *Based on the region-based similarity measure, a novel synthetic aperture radar (SAR) image retrieval method is proposed in this paper, which is inspired by the existing content-based image retrieval (CBIR) techniques and is oriented toward the Earth observation (EO). First, due to the large sizes of SAR images, new method semantically classifies the land covers in the patch level rather than the pixel level by the classic semi-supervised learning (SSL), which could reduce the workload of selecting the representative image patch and decrease the searching space in the similarity calculation component. Furthermore, to overcome the inevitable classification error, our method provides an error recovery scheme, preventing the errors produced in categorization to contaminate the retrieval results. Third, the similarity between two patches is calculated by the improved integrated region matching (IIRM) measure based on the region-based similarity measure, which fails to meet the expectation in SAR images. The proposed method can be embedded into any EO mining systems to help them complete the EO missions. After comparing the method presented in this paper to others, it is evident that our method performs more effectively than others from the CBIR aspect.*

Key Words: **Content-based image retrieval (CBIR), Earth observation (EO), improved integrated region matching (IIRM) measure, synthetic aperture radar (SAR) image retrieval, Semi-Supervised Learning (SSL).**

1. INTRODUCTION

Among those systems, Google's Image Search engine may be the most popular one at present. The truly successful general-purposed retrieval system has not emerged since the issues these systems address are too board, and the well-known semantic gap, i.e., the gap between low-level visual features and high-level semantics, exists in the *CBIR technique*. However, one of the extensions and utilizations of *CBIR* methods, RS image retrieval (RSIR), has been made greater success. Numbers of famous and practical *RSIR* methods have been proposed in recent years. Some of them focused on feature extraction and object semantic representation, and others were concerned on applying higher semantics.

There are numerous *RSIR* methods, while the methods aimed to general-purposed synthetic aperture radar (SAR) images retrieval are rare. As one of the EO products, SAR images have drawn increasing attentions recently due to the special characteristics such as possible of penetrating clouds, and operating in bad weather and night. With the increasingly SAR data open and free, a growing number of application on SAR data would be developed in future years. Thus, automatic and semiautomatic interpretation of these SAR data becomes both important and urgent, such as finding the interested regions of researchers from a large volume of images, understanding the distribution of land covers in one SAR image, and detecting changes of the same region in different times, etc.

In this paper, a novel SAR images retrieval method oriented toward the EO mining is introduced, which is based on region-based similarity measure. Different from many existing EO mining techniques, our retrieval method focuses on designing a robust similarity measure rather than using more complex information, such as metadata, to complete the mining mission. To construct the database, SAR images in big sizes are divided into equal sized image patches first, and then users pick up some representative patches. The semi-supervised learning (SSL) would be implemented to label all the image patches in database. Here, we adopt *SSL* to be classifier because it needs less train samples in categorization, which can ease the burden of the image patch selection. The classification step in our retrieval method could reduce the workload of the representative patch selection, narrow down the searching space in the resemblance calculation component, and decrease the impact of semantic gap to our method. When users input a query image patch, the similarities between the query and relevant target patches (the image patches exist in database) would be calculated by improved integrated region matching (IIRM) measure introduced in this paper for SAR images, and the retrieval results are displayed in order. The time consumption is concentrated on the similarity computation in our retrieval method. Due to the high efficiency of new similarity measure, and the pre -classification, our retrieval method is fast.

This paper focuses on the task of SAR image retrieval oriented toward the EO mining. Several improvements are added into traditional *CBIR* techniques to enhance the retrieval precision. The significant contribution of this work can be summarized as follows.

1. To narrow down the searching space in the similarity computation step and to enhance the retrieval speed, we add classification procedure into our method.

2. Considering the tough work of the representative samples selection, we adopt SSL to decrease the workload of choosing train set.

3. Due to the inevitable misclassification, a classification error recovery scheme is presented here to reduce the influence to the retrieval behavior.

4. With a view to the characteristics of SAR image, a new similarity measure named IIRM is introduced, and it is proved to be efficient and effective for SAR image based on the experiment.

2. METHODOLOGY

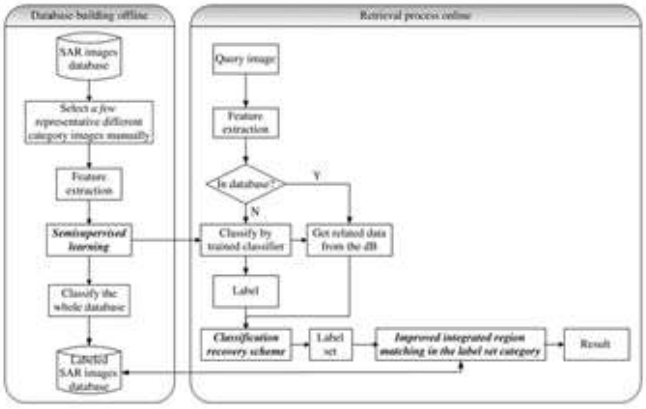

Fig -1: Framework for SAR Image Retrieval Method

This paper is organized as follows: In Section II, the framework of SAR retrieval method is described. Section III introduces the proposed SAR retrieval approach in detail. Section IV discusses our experiments, and Section V gives the conclusion and suggests the future work.

2.1 FRAMEWORK OF SAR RETRIEVAL METHOD

The framework of the proposed SAR retrieval approach is shown in Fig. 1, arising from the content-based SAR image retrieval approach. This framework can be separated into two parts, including "database building" and "retrieval process." For a query image patch, the retrieval process begins with feature extraction. The features of target patches are usually pre calculated and stored in the "database building" part. With these features, the semantic category of the query image patch, and the resemblance between the query and relevance category target patches, can be calculated. Finally, the retrieval results are listed in order. Pre classification is used to enhance the retrieval precision and reduce the matching space in database. Below we discuss each component of the proposed method.

2.1.1 Database Building Offline

The primary target of this part is to build a labeled SAR images database, including initial images acquisition, representative patches selection, feature extraction and database categorization. Suppose that there is a set of M raw SAR images I_i,

i = 1, . . . , M of any type (e.g., Radar Sat and Terra SAR). First, we divide them into equal sized non overlapping patches of size $X \times Y$, getting N patches $\{p_1, p_2, \ldots, p_N\}$. Then the features for categorization of all patches are extracted. After that, a certain number of representative patches are selected and labeled manually to train the classifier. Finally, image patches in the database are labeled by the classifier trained before.

Since the selection of representative images is a time consuming and human sensitive work generally, the number of them should be as small as possible. To solve the issue mentioned above, *SSL* is adopted. The *SSL* can be regarded as the supervised learning with additional information on the distribution of the data, and it will be useful whenever there are far more unlabeled data than labeled in practice. To choose representative samples from database with huge number of image patches, we first use the *k*-means algorithm for clustering all of patches. It turns out that *k*-means cluster centers have a stronger representation power to adequately cover the vast samples. The *k*-means algorithm does not specify the number of clusters, so we should determine the number of clusters depending on the number semantic categories. In this paper, the number of clusters is 20. After the clustering, the range of representative sample candidates is reduced, i.e., the patches which are near to the cluster centers. Then we only select the representative samples from these candidates. In this work, the number of candidates and representative samples are 15% and 5% of total number of image patches.

2.1.2 Mining Process Online

The major objective of this part is searching the most similar image patches to a query from the labeled database. When a query patch is input, its feature is extracted immediately with the same method as the online process. Then two cases should be considered, that is, whether the patch exists in the labeled image database or not. If the query patch exists in the database, we can get the relevant knowledge from the database directly. Otherwise, the query patch should be categorized by the *SSL* method first, and then the similarities between the query and its relevant image patches are calculated. Considering the characteristic of SAR image, a new similarity measure is proposed in this paper based on the integrated region matching (IRM) mea-sure , named *improved integrated region matching*. Here, "its relevant image patches" means that we do not only match the patches belonging to the same category with the query but also the patches belonging to the relevant category, to overcome the possible classification error. The definition of "relevant

category" is presented in Section III-B, which is obtained through the confusion matrix.

Three notable improvements are obtained by the proposed retrieval method. First, the workload of picking representative image patches is reduced by the *SSL* method. Second, the empirical confusion matrix is used to decrease the effect resulted from inevitable classification error. Third, a new similarity metric is proposed specially for SAR images, based on the IRM measure. With the help of the tips mentioned above, the mining precision is enhanced significantly.

3. CONCLUSIONS

Inspired by the CBIR techniques and its application on EO mining, we propose a novel retrieval method for SAR images, one of the EO products, based on semantic classification and region-based similarity measure. On the basis of classical similarity measure IRM, we design the SAR-oriented similarity measure *IIRM*. The encouraging experiment results show the performance of *IIRM* is suitable well to SAR images. Moreover, semantic categorization via *GSSL* not only reduces the work-load of train set selection manually, but also narrows down the searching space in database. In addition, this scheme can decrease the impact of semantic gap to our retrieval method. Considering the inevitable classification error, we provide the *error recovery scheme* by the empirical confusion matrix, which prevents the errors produced in categorization to contaminate the retrieval results. By our retrieval approach, users can understand several SAR images simultaneously rather than a single one. Numerical assessments on the dataset indicate the good quality of our new SAR-oriented retrieval scheme.

The proposed system tries to fulfil the following objectives:

1] To narrow down the searching space in the similarity computation step and to enhance the retrieval speed, we add classification procedure into our method.

2] Considering the tough work of the representative samples selection, we adopt SSL to decrease the workload of choosing train set.

3] Due to the inevitable misclassification, a classification error recovery scheme is presented here to reduce the influence to the retrieval behavior.

4] With a view to the characteristics of SAR image, a new similarity measure named IIRM is introduced and it is proved to be efficient and effective for SAR image based on the experiment.

REFERENCES

[1] Licheng Jiao, "SAR Images Retrieval Based on Region-Based Similarity Measure for Earth Observation" IEEE journal of selected topics in applied earth observations and remote sensing, VOL. 8, NO. 8, AUGUST 2015.

[2] Daniela Espinoza-Molina, "Earth-Observation Image Retrieval Based on Content, Semantics, and Metadata" Manuscript received September 28, 2012; revised January 11, 2013 and March 20, 2013; accepted April 15, 2013.

[3] Xu Tang, "SAR Image Content Retrieval Based on Fuzzy Similarity and Relevance Feedback" Manuscript received May 10, 2016; revised October 19, 2016 and January 22, 2017; accepted February 1, 2017

[4] Xu Tang," Fusion Similarity-Based Re-ranking for SAR Image Retrieval" Manuscript received July 4, 2016; revised November 25, 2016; accepted December 3, 2016.