

EVALUATING THE MEASUREMENT SCALES OF SEMANTIC FEATURES FOR REMOTE SENSING IMAGES RETRIEVAL

Changxin Gao, Nong Sang*, Qiling Tang

Institute for Pattern Recognition and Artificial Intelligence,
National Key Laboratory of Science & Technology on Multi-spectral Information Processing,
Huazhong University of Science and Technology, Wuhan, 430074, China

changxin.gao@gmail.com nsang@hust.edu.cn tqlinn@sohu.com

Commission I, The Role of Geomatics in Decision-Making

KEY WORDS: Image retrieval, Semantic features, Content-based image retrieval, Measurement scale, ordinal scale

ABSTRACT:

This paper presents a method for retrieval of remote sensing images using low dimensional code of the spatial layout of a scene, which is proposed to represent the meaning of the scene. This semantic feature is based on the statistical features of gradient orientations. To capture coarse spatial information, the semantic feature is represented in a pyramidal structure. The level of the pyramid structure is evaluated in this paper for the remote sensing image retrieval tasks. Many features are proposed to describe the semantic information of the images, however, the scale of the measurement of them are no or less discussed in these researches. There are three candidates of measurement scale for image retrieval task: the interval scale, ordinal scale, and ratio scale. To evaluate the three measurement scales of semantic features for remote sensing image retrieval, this paper build a remote sensing image dataset, which is composed of ten image categories: Olympic gymnasium, urban area, campus of university, island, navy base, aircraft carrier, nuclear reactor, air base, cloverleaf junction, and mountain peak. The average normalized modified retrieval rank (ANMRR) are used to estimate the performance of semantic features with different measurement scales for remote sensing image retrieval task on our dataset. These experimental results demonstrate that ordinal scale is more effective for image retrieval task.

1. INTRODUCTION

Current remote sensing platforms for Earth observation together acquire several terabytes of image data per day. The storage, organization and retrieval of the image data pose a challenge and important problem. And, due to this data volume the storage, organization and retrieval of these images became increasingly complex. According to the different organization and management schemes for Remote Sensing image data, the main methods used for remote sensing image data retrieval at present include: (1) File-based image data retrieval; (2) Metadata-based retrieval of mapsheet; (3) Image content-based semantic retrieval (Li and Bretschneider, 2004).

Content-based image retrieval (CBIR) technology is an image retrieval technology using image vision contents such as color, texture, shape, spatial relationship, not using image notation to search images (Manjunath and Ma, 1996; Ngo et al, 2002; Li and Bretschneider, 2004; Li and Bretschneider, 2006). CBIR techniques together with the query-by-example approach enable a powerful and straightforward procurement of remote sensing images in the immense database. In order to process queries directly related to content information of remote sensing images, a variety of content-based image retrieval techniques were applied based on abstract automatically extracted features.

Although varying success was reported, the main difficulty is that the retrieval performance is limited by the semantic gap

between low-level features and high-level semantic concepts (Lee et al., 1999; Rui et al., 1998). In order to bridge the semantic gap, a scene context feature, which performs well in nature scene categorization, is applied to describe the meaning of the remote sensing images and to perform CBIR tasks.

Traditional conceptions of research in computer vision treat objects as the atoms of recognition. In contrast, some experimental studies have suggested that the recognition of real world scenes may be initiated from the encoding of the global configuration, without necessarily identifying the objects they contain (Biederman et al., 1982; Schyns and Oliva, 1994; Oliva and Schyns, 2000; Greene and Oliva, 2005). Human are able to comprehend the amount of perceptual and semantic information (the semantic category of each image as well as a few objects and their attributes) within a single glance in less than 200 ms, which refers to gist of the scene (Potter, 1975; Henderson and Hollingworth, 1999; Oliva, 2005). In recent works, low dimensional code of the spatial layout of a scene is proposed to represent the characteristics of a scene (Murphy et al., 2003), which yield a rich set of cues to its semantic category.

Hence, this paper presents a method for retrieval of remote sensing images using low dimensional code of the spatial layout of a scene, which is proposed to represent the meaning of the scene. In this work, the semantic feature is based on the statistical features of gradient orientations. To capture coarse spatial information, the semantic feature is represented in a

* Corresponding author.

pyramidal structure. The level of the pyramid structure is evaluated in this paper for the remote sensing image retrieval tasks.

Many features are proposed to describe the semantic information of the images, however, the scale of the measurement of them are no or less discussed in these researches. There are three candidates of measurement scale for image retrieval task: the interval scale, ordinal scale, and ratio scale. Most traditional image representation approaches apply interval scale; however, recent researches suggest that ordinal scale performs well for image representation. This paper evaluates the performance of our semantic feature with three different scales of measurement, for the remote sensing image retrieval tasks.

To evaluate the three measurement scales of semantic features for remote sensing image retrieval, we build a remote sensing image dataset, which is composed of ten image categories: Olympic gymnasium, urban area, campus of university, island, navy base, aircraft carrier, nuclear reactor, air base, cloverleaf junction', and mountain peak. Each category has more than 30 images, and the image size is 1148×793 pixels. The sources of the pictures in the dataset is Google Earth, all of them are cropped from Google Earth. The average normalized modified retrieval rank (ANMRR) are used to estimate the performance of semantic features with different measurement scales for remote sensing image retrieval task on our dataset. The results show that ordinal scale is more effective.

2. SCALES OF MEASUREMENT

To analyze data more effectively, the scales of measurement should be considered. Stevens's theory (Stevens, 1946) of scales of measurement has been an important methodological resource within psychology for half a century. It advanced the representational theory of measurement and promised to open up to scientific investigation the issue of the structure of psychological attributes. And this paper follows these concepts. Stevens presented four types of scales of measurement: nominal scale, ordinal scale, interval scale, and ratio scale.

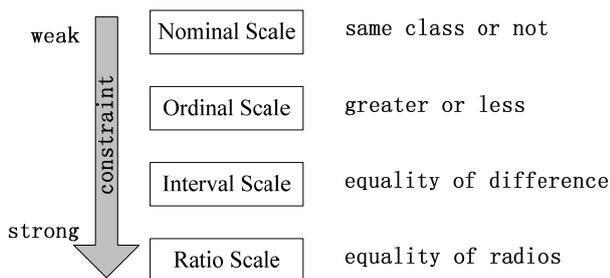


Figure 1. The properties and relationships of the four scales

The meanings of four types of scales are much different. The nominal scale represents the most unrestricted assignment of numerals, which remains invariant under general substitution or permutation group. The ordinal scale comes from the operation of rank ordering and assigns values to objects based on their ranking with respect to one another, which. The interval scale represents a quantitative measurement, which remains invariant when a constant is added. The ratio scale is the most restricted measurement. The properties and relationships of the four scales are shown in Fig. 1.

For image retrieval tasks, the nominal scale is really a list of categories to which objects can be classified. Thus, it can not be used in feature representation. This paper evaluates the performance of the four scales for image retrieval.

3. FEATURE EXTRACTION

Traditional conceptions of research in computer vision treat objects as the atoms of recognition. In contrast, some experimental studies have suggested that the recognition of real world scenes may be initiated from the encoding of the global configuration, without necessarily identifying the objects they contain (Biederman et al., 1982; Schyns and Oliva, 1994; Oliva and Schyns, 2000; Greene and Oliva, 2005). Some experiments in perception and visual search suggest that the visual system first processes context information in order to index object properties. This paper extracts scene context features for image retrieval, which uses pyramidal histograms of gradient orientations information.

3.1 Feature with different measurement scale

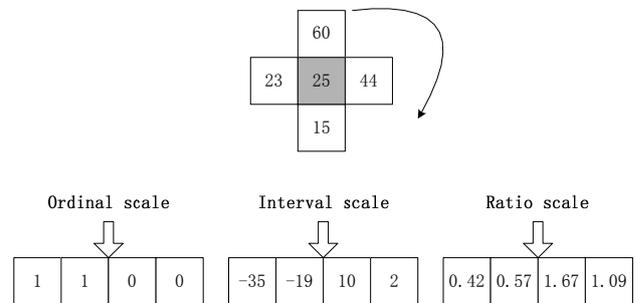


Figure 2. Illustration of the features extraction with different levels.

This section presents the comparison of feature extraction with different levels: ordinal scale, interval scale, and ratio scale. Each position is measured with its neighbours, and the four-neighbour case is shown in Fig. 2. It is noted that the ordinal representation is similar as LBP (Ojala et al., 1996). Also noted that the constraint of ordinal scale is weakest, and which of ratio scale is strongest.

3.2 Pyramidal structure

To obtain coarse spatial information, we employ pyramidal structure to extract features at different scales. A three-level pyramidal structure is shown in Fig. 3. Extracting pyramidal histogram features consists of three stages: first we build the pyramidal structure by dividing one region to four sub-regions; then we compute the histograms of all the regions (or sub-regions); finally, we get the representation vector by combining all the histograms together. These global features are hierarchical representation, the levels of which may be used to balance the trade-off between generalization ability and discrimination ability. The discrimination ability is improved while the generalization ability is weakened when the levels increase, and vice versa.

4. EXPERIMENTS

In this section, we report comparison results of image representation with different scales for remote sensing images

retrieval task. The number of levels of pyramidal structure is also evaluated.

4.1 Dataset

To evaluate the three measurement scales of semantic features for remote sensing image retrieval, we build a remote sensing image dataset, which is composed of ten image categories (semantic concepts): Olympic gymnasium, urban area, campus of university, island, navy base, aircraft carrier, nuclear reactor, air base, cloverleaf junction, and mountain peak. Each category has more than 30 images, and the image size is 1148×793 pixels. The sources of the pictures in the dataset is Google Earth, all of them are cropped from Google Earth. Some examples are shown in Fig. 4. Some categories are easily confused, for example, urban area & campus of university, and navy base & aircraft carrier.

4.2 Experimental results

For our experiments, we perform remote sensing image retrieval task, that is to say, retrieving images from a set of images by a query example (image), called query-by-example task. Comparison results of image representation with different scales for remote sensing images retrieval task are presented. The average normalized modified retrieval rank (ANMRR) (Lee et al., 2003) are used to estimate the performance of semantic features with different measurement scales for remote sensing image retrieval task on our dataset. Both performance measures are popularly used in the current researches. The values of ANMRR range between $[0, 1]$. A low value of ANMRR indicates the high retrieval rate with relevant shots ranked at the top.

ANMRR: Let Q as the number of queries and N as the number of items in a database. For a query q , $R(q)$ is defined as the set of relevant items in a dataset for q , and $NR(q)$ as the number of items in $R(q)$. Then, ANMRR is computed as

$$ANMRR = \frac{1}{Q} \sum_{q=1}^Q \frac{MRR(q)}{K + 0.5 - 0.5 \times NR(q)}$$

where $K(q) = \min(4 \times NR(q), 2 \times t)$, $t = \max_{k=1}^Q (NR(k))$

$$MRR(q) = \frac{1}{NR(q)} \sum_{k=1}^N \text{Rank}(k, q) - 0.5 - \frac{NR(q)}{2}$$

The function $\text{Rank}(k, q)$ computes a value for an item which is retrieved as the k th most similar item to query q as

$$\text{Rank}(k, q) = \begin{cases} k & \text{if } k \leq K(q) \text{ and } k\text{th item} \in R(q) \\ K(q) + 1 & \text{if } k > K(q) \text{ and } k\text{th item} \in R(q) \\ 0 & \text{otherwise} \end{cases}$$

Table 1. The ANMRR with different measurement scales and pyramid levels. The highest result is shown in bold.

ANMRR	Scale		
	Ordinal	Interval	Ratio
1-level	0.5143	0.5360	0.5324
2-level	0.4747	0.5204	0.5745
3-level	0.4730	0.4849	0.6042
4-level	0.4731	0.4870	0.6224

The ANMRR results with different measurement scales and pyramid levels are shown in Table 1. It can be seen that: (1) the

ordinal scale performs best, while the ratio scale performs worst; (2) the performance the representation with 3-level pyramidal structure is best, e.g., which achieves the best balance between generalization ability and discriminative ability. Fig. 5, Fig. 6, and Fig. 7 show the query results with the ordinal scale, the interval scale, and the ratio scale, first 20 images with lowest distances are given for each query (3-level pyramidal structure).

Our experiments demonstrate that semantic feature of remote sensing images are surprisingly effective for image retrieval. These results also give the guide of model selection in this representation, three-level pyramid structure and ordinal scale should be employed for remote sensing image retrieval.

5. CONCLUSIONS

Image representation is very important for remote sensing retrieval; however, model selection is not resolved. This paper evaluates the performance of semantic features with different measurement scales and the pyramid levels for remote sensing image retrieval. To analyze the performance, we build a remote sensing image dataset, which is composed of ten semantic categories. Our experimental results show that ordinal measurement scale and 3-level pyramid structure is more effective for image retrieval.

6. ACKNOWLEDGEMENTS

This work is supported by the Project of the National Natural Science Foundation of China under Grant No.60736010, the 863 Project of China under Grant No.2007AA01Z166, and the Program for New Century Excellent Talents in University under Grant No. NNCET-05-0641.

REFERENCES

- Biederman, I., Mezzanotte, R. J., and Rabinowitz, J. C., 1982. Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology*, vol.14, pp. 143-177.
- Greene, M. R., and Oliva, A., 2005. Better to run than hide: the time course of naturalistic scene decisions [Abstract]. *Journal of Vision*, 5(8), pp.70a.
- Henderson, J. M., and Hollingworth, A., 1999. High level scene perception. *Annual Review of Psychology*, vol. 50, pp. 243-271.
- Lee, C., Ma, W. Y., and Zhang, H. J., 1999. Information embedding based on user's relevance feedback for image retrieval. In: *Proc. SPIE Conf. Multimedia Storage and Arch. Syst.*, vol. 5, pp. 20-22.
- Lee H.Y., Lee H.K., and Ha Y.H., 2003. Spatial colour descriptor for image retrieval and video segmentation. *IEEE Trans. on Multimedia*, 5(3), pp. 358-367.
- Li, Y., and Bretschneider, T., 2004. Semantics-based satellite image retrieval using low-level features. *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 7, pp. 4406-4409.
- Li, Y., and Bretschneider, T., 2006. Remote sensing image retrieval using a context-sensitive Bayesian network with

relevance feedback. *International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 2461-2464.

Manjunath, B.S., and Ma, W.Y., 1996. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8), pp. 837-842.

Murphy, K. P., Torralba, A., and Freeman, W. T., 2003. Using the forest to see the trees: a graphical model relating features, objects and scenes. *NIPS*, vol. 16, pp. 1499-1507.

Ngo, C.W., Pong, T.C., and Zhang H.J., 2002. On clustering and retrieval of video shots through temporal slices analysis. *IEEE Transactions on Multimedia*, 4(4), pp. 446-58.

Ojala, T., Pietikainen, M., and Harwood, D., 1996. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1), pp. 51-59.

Oliva, A., 2005. Gist of the scene, In *Neurobiology of Attention*, L. Itti, G. Rees and J. K. Tsotsos (Eds.), Elsevier, San Diego, CA, pp. 251-256.

Oliva, A., and Schyns, P. G., 2000. Diagnostic colors mediate scene recognition. *Cognition Psychology*, vol. 41, pp. 176-210.

Potter, M. C., 1975. Meaning in visual search. *Science*, vol. 187, pp. 965-966.

Rui, Y., Huang, T. S., Ortega, M., and Mehrota, S., 1998. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Trans. Circuits Syst. Video Technol.*, 8(5), pp. 644-655.

Schyns, P. G., and Oliva, A., 1994. From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition. *Psychology Science*, vol. 5, pp. 195-200.

Stevens, S.S., 1946. On the theory of scales of measurement, *Science*, 103(2684), pp. 677-680.

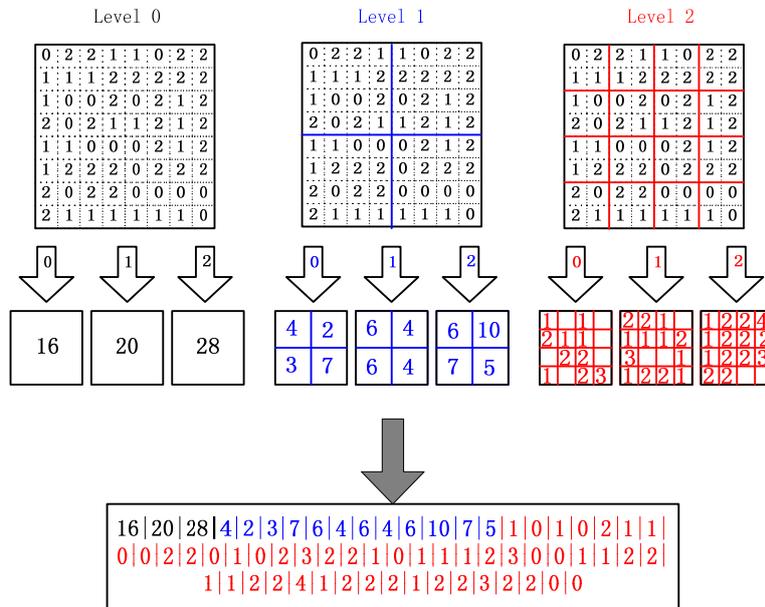


Figure 3. Illustration of coarse spatial relationship extraction with the pyramid structure.

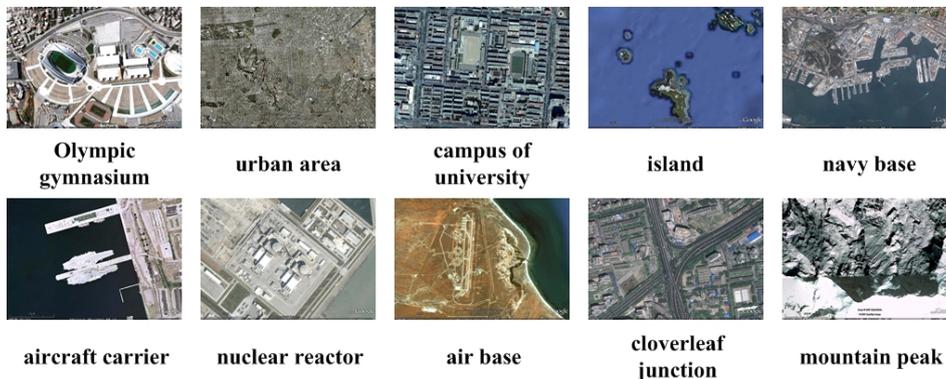


Figure 4. Some examples our dataset.



Figure 5. Query results with ordinal measurement scale.

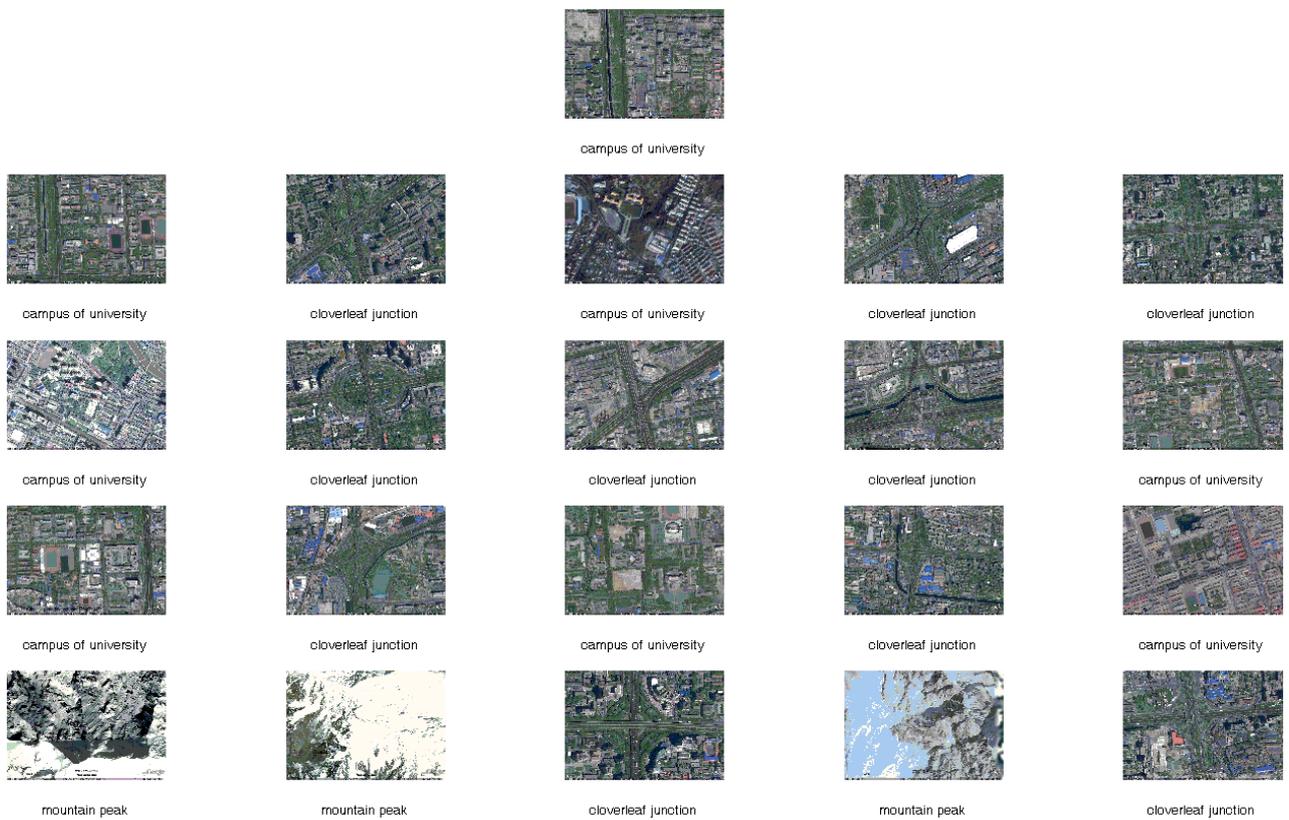


Figure 6. Query results with interval measurement scale.

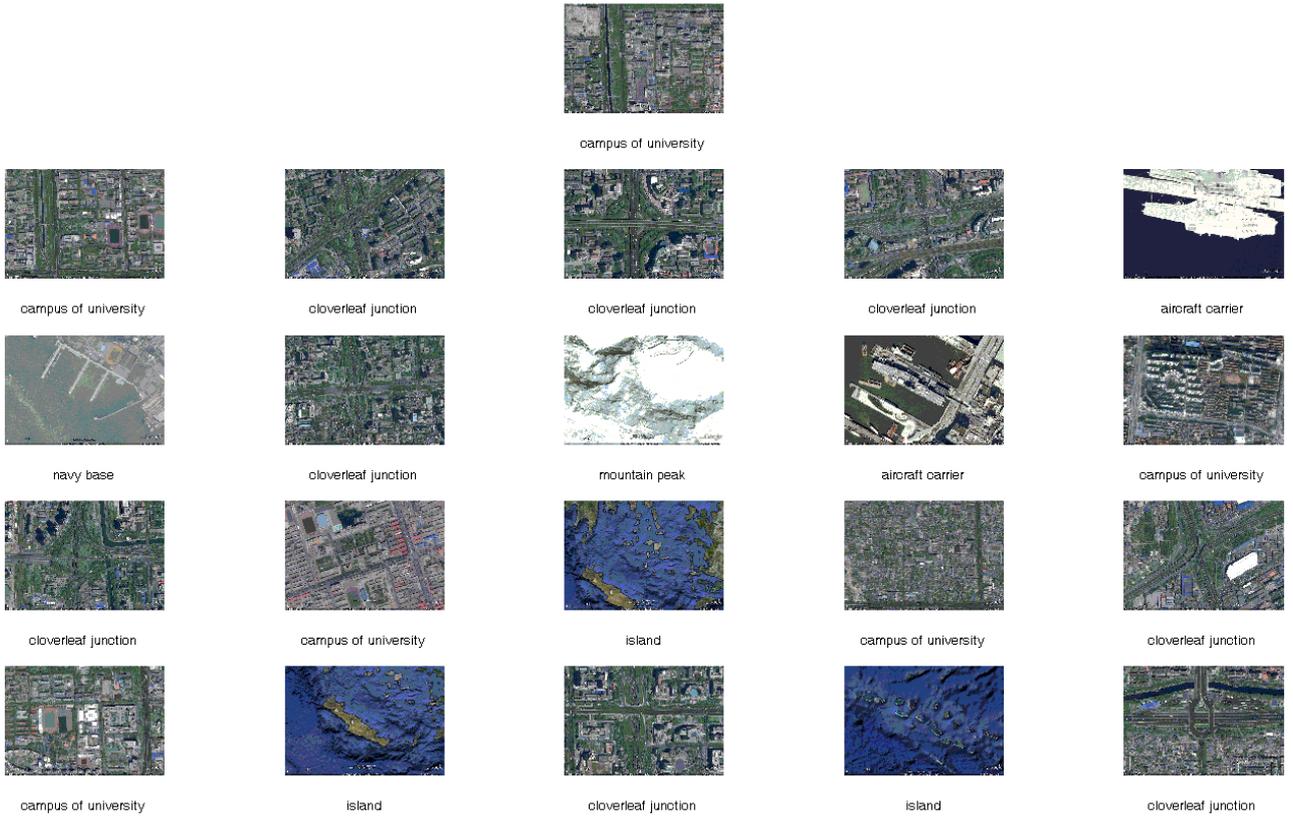


Figure 7. Query results with ratio measurement scale.