58th ILMENAU SCIENTIFIC COLLOQUIUM
Technische Universität Ilmenau, 08 – 12 September 2014
URN: urn:nbn:de:gbv:ilm1-2014iwk:3

GLOBAL DESCRIPPTORS APPLICATION IN OBJECT RECOGNITION

Dmitry Kupriyanov[1], Dmitry Shvarts[2], Victor Musalimov[1], Mart Tamre[2]

1-ITMO University, 2-Tallinn University of Technology

ABSTRACT

Nowadays problem of group robot interaction is as actual as it never was before. SLAM technology is a step forward in solving it. You can create large map with the help of the group of mobile robots by merging pieces of navigational data that robots collected. It would be nice to make robots also able to help each other determine precise size and position of object.
The work deals with series of experiments, which include image recognition and retrieval of corresponding object from database and so on. Local descriptors approach are compared to global one, so usage of SIFT (Scale-invariant feature transform) and GIST feature descriptors is researched.

Index Terms – computer vision, GIST, SIFT

1. INTRODUCTION

Current development of computer vision technologies made it possible to implement them widely in SLAM systems. It can be very beneficial, especially if applied to robots operating in non-prepared environment. Vision feature descriptors can lend a helping hand in making truly autonomous robots, independent from environment where they are intended to operate. Modern vision feature systems with high accuracy usually use different vision feature descriptors and filtering algorithms. These systems can be very demanding in means of calculation power, which is inappropriate when applied to small mobile robots or servers controlling large robotic hives.

Figure 1. The colony of warehouse robots

2. SINGLE FEATURE APPROACH

In SLAM we can try to use single feature classifier, which will reduce accuracy in comparison with multi-feature systems, but by imprinting navigation data into each image descriptor we can make this accuracy loss not that drastic. This approach should be the best way of SLAM realization for large groups of robots. Current feature detection algorithms usually use several different feature descriptors and more adapted for analysis of large groups of images, due to that they are pretty heavy to use in robotics. These combined classifiers cn be very helpful in laboratories, where researchers have to deal with large amounts of visual data [1] or for processing large image databases [2].

Global features have been used in the computer vision community as an alternative to local features for scene classification [3], [4]. Their key advantage for this application is that their performance is very similar to that of local features at a much lower cost [5]. In recent years, in view of global features, greatest interest has been shown in the robotics community. Most shape and texture descriptors may belong to this category. Such features are attractive because they produce very compact representations of images, where each image corresponds to a point in a high dimensional feature space. Global descriptor may be successfully applied for scene classifications represented on the base level or on the subordinate level of scene description. The results of experiments confirm the validity of this claim. Query image represented as a high-dimensional vector is compared against all images from the database represented in the form of high-dimensional vectors too. In the experiment and in the further work, this thesis uses the popular GIST descriptor. Minimum Euclidean distance between the compared GIST descriptors, meaning the query image and its likely candidate, determines the degree of similarity of images.

Global feature descriptors can be successfully implemented in SLAM systems, for example, GIST classifier can be used to create and merge 3D maps, created by groups of SLAM robots [6].

If we continue following the optimization path, it would seem logical to use global feature descriptor instead of local one. Global descriptors work much faster than local ones, because they describe image as a whole instead of searching objects and determining areas of interest. Main goal of current research was to compare accuracy of two image classifiers, based on local and global feature descriptors.

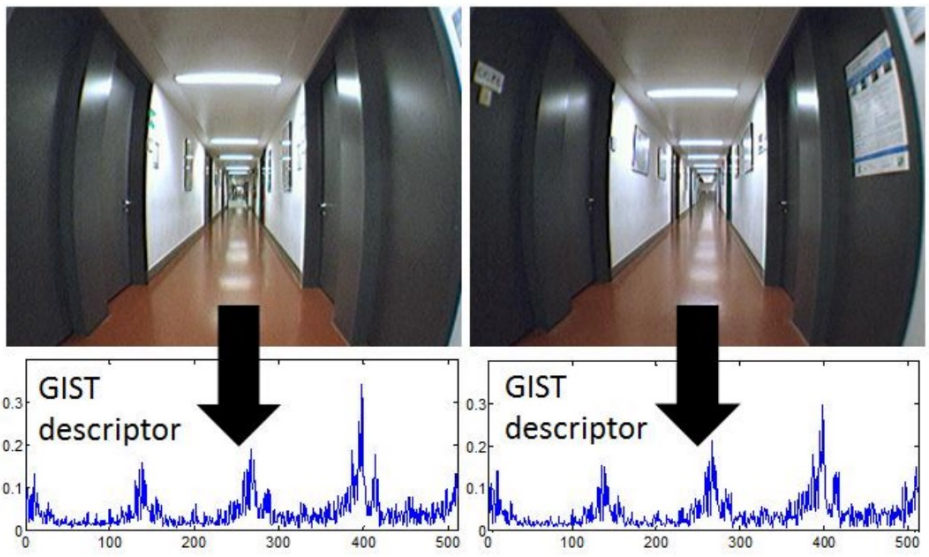

Figure 2. When different places look similar they result in similar features, that is where positioning is needed.

3. EXPERIMENT

In our experiment we compared classifiers based on Dense SIFT from VLFeat library and GIST. GIST classifier used images prefiltered with Gabor filter.
In the beginning there was rather small database of 90 images divided into three groups: flat object (keycard), simple 3D object (black box), and more complicated 3D object (mobile robot tracks). Each of the objects was represented by 30 images. Such small database was used to imitate the situation, when robot only starts movement and has no data about its surroundings.

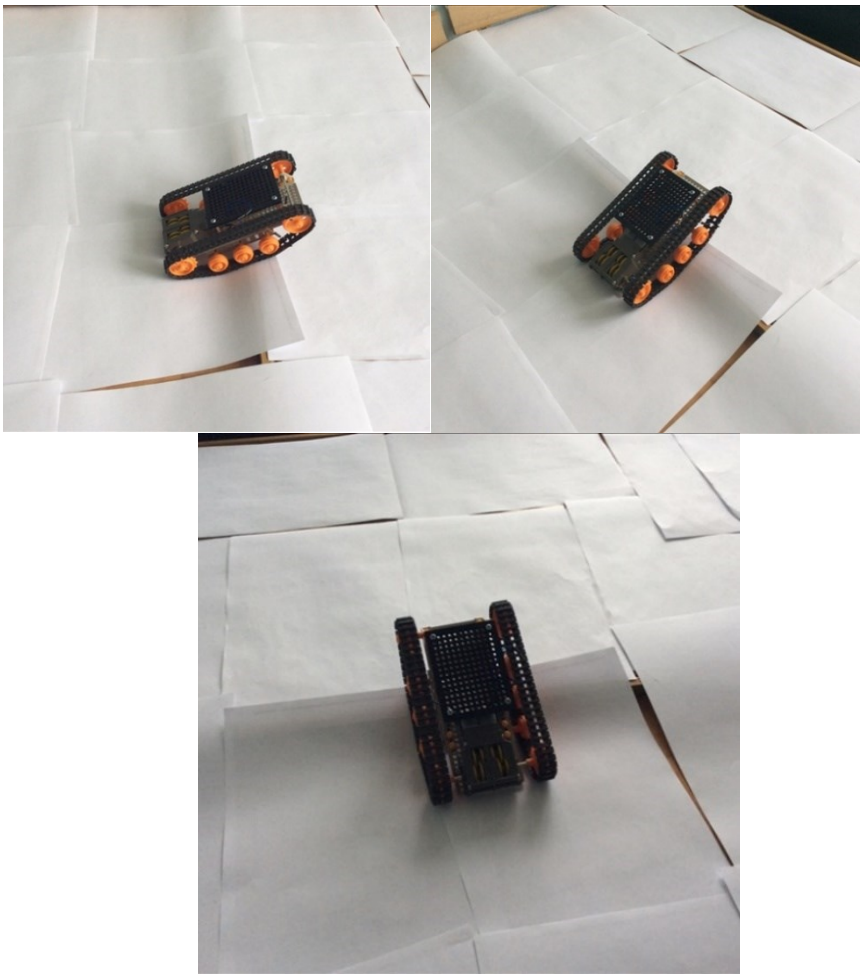

Figure 3. Some of the photographs used for test

Classifiers were built using above named visual features. Each of classifiers used 15 images for training per group and compared them to all remaining 75 images. 15 images is the approximate number of features which robot will produce in 10 seconds after turning on.
In GIST classifier we used spatial envelope, procedure based on a very low dimensional representation of the scene, as our main intention was to enhance visual SLAM map without making algorithm more complicated.

4. RESULTS OF EXPERIMENT

After creating equal conditions for each of classifier, we received really different results. This can be explained by the difference in nature of local and global descriptors. It can be seen that

local image features returned better results dealing with complex 3D structure where points of interest can be found easily. On the contrary, global features were more effective when describing an object with large and contrast borders. Keycard, as an object combining both properties of box and cart was identified with practically the same results by both descriptors. GIST classifier showed higher overall accuracy. This can be explained by unified background of all images.

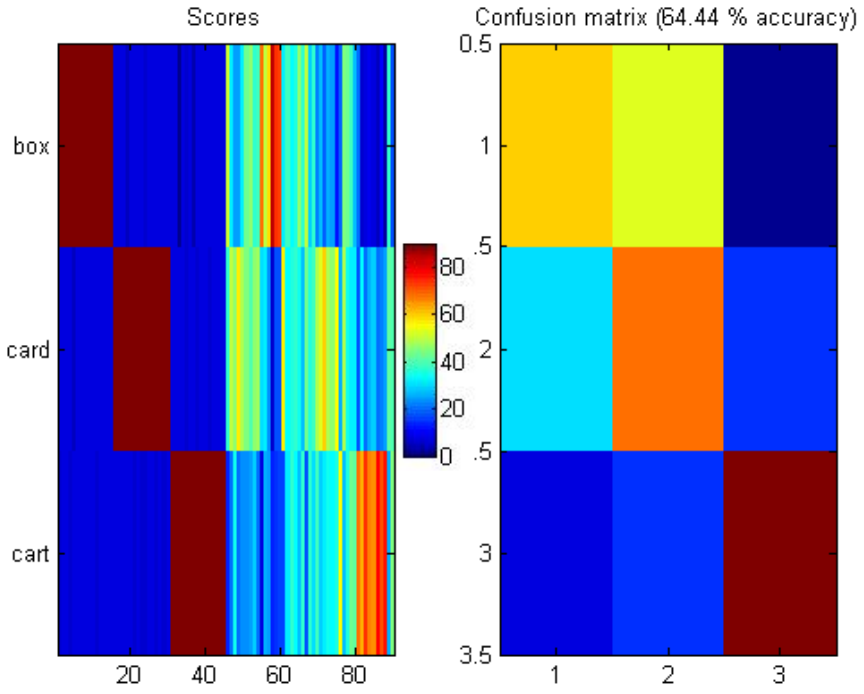

Figure 4. SIFT classifier confusion matrix

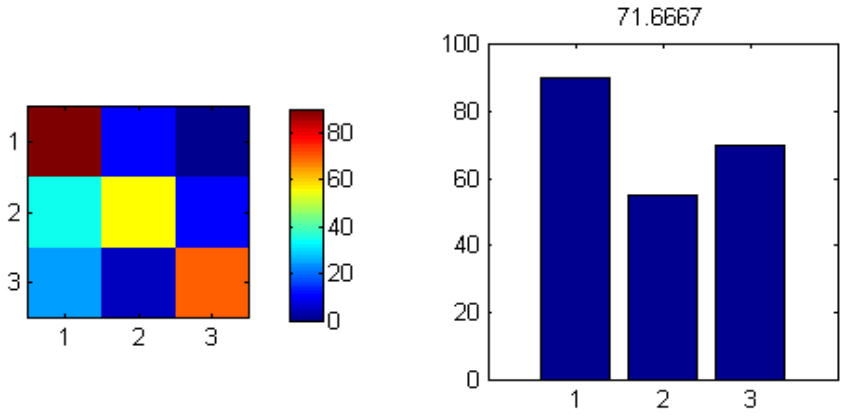

Figure 5. GIST classifier confusion matrix (1 – glasses case, 2 – mobile robot tracks, 3 – card)

Figure 4 represents results of SIFT classifier. On the left side we can see probabilities of assigning each picture to one of three groups. Dark red regions represent images used for training. To the right is confusion matrix where overall accuracy for the whole group of images is depicted. It is obvious that classifier was confused about simple objects to, while having almost 100% accuracy in determining complex structure of track cart.
Figure 5 shows results of GIST classifier. On the left of figure there is a confusion matrix, while on the right histogram with overall group accuracy is presented. Overall accuracy is 7% higher, but the best results were shown while determining black glasses case.

5. CONCLUSIONS

This small experiment showed that while robot operates in environment with unified background global feature descriptors could be even more accurate than the local ones. Of course, there are several restrictions. The whole principle of global feature description suites best for working with contrast objects, while it has poor results working with complex structures. In future, global description algorithm with adjustable frames can be very useful in SLAM systems, as it could give object detection systems a boost.
There are actually two ways of developing global feature based object recognition. One, as was mentioned above, includes modifying description algorithm, for example, with adding adjustable frames. This can be difficult and, what's more, there is a chance that this will make global feature descriptors too complicated, so there will be no benefits of using them instead local ones.
The other way is to limit possible areas where such recognition can be applied. It is possible to determine groups of objects, which can be easily recognized by global feature based object recognition systems. It won't make such systems truly universal, but they can become simple and low cost solution in some fields, for example, in automatized warehouses. A colony of industrial robot platforms, without any central control system able to perform logistic tasks in autonomous mode together with the system's ability to automatically adapt to any environment, make it available for both large and small warehouses. This way of development seems more appropriate.

REFERENCES

[1] D.A. Lisin, M.A. Mattar, M.B. Blaschko, M.C. Benfield and E.G. Learned-Miller, "Combining Local and Global Image Features for Object Class Recognition," Vision Lab, University of Massachusetts, Amherst, pp. 1-10, 2005.
[2] K. Murphy, A. Torralba, D. Eaton and W. Freeman, "Object Detection and Localization Using Local and Global Features", CSAIL, MIT, Cambridge, 2005.
[3] A. Oliva and A. Torralba . Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. Int. J. Comput. Vision, vol. 42, number 3, pp.145-175, 2001
[4] V. Aditya, J. Anil and Z. Hong Jiang . On image classification: city images vs. landscapes. Pattern Recognition. vol. 31, pp. 1921-1935, 1998
[5] M. Douze, H. Jégou, H. Sandhawalia, L. Amsaleg and C. Schmid. Evaluation of GIST descriptors for web-scale image search. Proceedings of the ACM International Conference on Image and Video Retrieval, Article 19, 8 pages, 2009
[6] D. Shvarts, Global 3D Map Merging Methods for Robot Navigation, TUT Press, Tallinn, 2013

CONTACTS

D. Kupriyanov	qudmv@yandex.ru
Prof. V. Musalimov	musvm@yandex.ru
PhD. D. Shvarts	shvarts.dmitry@gmail.com
Prof. M. Tamre	mart.tamre@ttu.ee