Combining Color and Haar Wavelet Responses for Aerial Image Classification

Ricardo C. B. Rodrigues[1], Sergio Pellegrino[1], and Hemerson Pistori[2]

[1] Instituto Tecnológico de Aeronáutica
P. Marechal Eduardo Gomes, 50 Sao Jose dos Campos, SP, Brazil
http://www.ita.br
[2] INOVISAO, Universidade Católica Dom Bosco
Av. Tamandaré 6000, Campo Grande, MS 79117-900, Brazil
{rcezar,pell}@ita.br
pistori@ucdb.br

Abstract. A new set of attributes combining color and SURF-based histograms coupled with a SVM classifier to enhance visual based autonomous aerial navigation is proposed. These new features are used for region classification with aerial images in order to speed up the UAV (Unmanned Aerial Vehicles) localization performed by image matching using only reference images according to the region classification. Experimental results comparing the proposal with color or SURF only attributes are presented. In the experiments the UAV localization task can be performed four times faster using the proposed approach, however the performance gain can be still bigger for large datasets of reference images.

Keywords: unmaned aerial vehicles; vision based navigation; feature extraction; SURF

1 Introduction

During the last two decades the interest in UAVs had an exponential growth. They became an essential weapon in military fields especially after 2001 with the successful record of North American model MQ-1 Predator in missions in Pakistan, Uzbekistan and Afghanistan [1]. With the development of small and mature applicable technologies such as digital cameras and GPS (Global Positioning System) UAVs have also emerged as solution for many civil applications such surveillance, firefighting, remote sensing, etc [2].

Currently there is a large variety of UAV models with different sizes, shapes and characteristics, such diversity boosted the development of UAV flight hardware [2]. But what remains as a challenge is the study on how to provide autonomy to UAV's and what degree of autonomy can be reached. In this sense, image-based robots navigation has been a subject investigated by many research groups and some works have exploited the supervised approaches to perform unmanned aerial vehicle (UAV) autonomous navigation [3–6].

One important computer vision supervised approach for robots navigation is the image registration, a supervised method which register target images in a database and match them with current images during navigation. For each match performed, the robot can use georeferences associated to images in database (knowledge base) and then estimate robot position. In [6] Conte uses the image registration approach to update the accumulated error of a visual odometer. So whenever a match is reliable it restores absolute UAV position and the filter is updated. The result of this work shows that this technique can reproduce a similar path to the GPS navigation system. But the image matching used is based on border filters that may be not so effective for some variances in scales, perspective or other changes in the environment.

Today there are many available databases of georeferenced aerial images, also called waypoint images or simply waypoints, which could be used to aid in UAV's autonomous navigation. The problem is how to match current images with database considering image variances in scale, rotation, perspective illumination and so on. The work [7] presents an evaluation of a supervised method based on image registration approach using robust scale-invariant algorithms like SIFT and SURF for waypoint recognition. The results demonstrate that these algorithms are able to accurately match the images, however this approach could be unfeasible to real time aircrafts navigation due to high processing time required for image matching using large databases.

In this context this work proposes a coarse geografical region classification step using color and SURF based descriptors. In this way, the search for waypoints using image match will be bounded to the images in the dataset that are related only to one region previously found and not to the entire waypoints dataset. The next section presents the proposed approach and is followed by the experiments section which reports the evaluation of three different classifiers and three sets of image descriptors. Conclusions and future works are presented in the last section.

2 Background

The scale-invariant SURF algorithm is used to detect interest points, named here keypoints, and describe local features of the images. This algorithm was chosen due to its performance and accuracy in detecting and matching keypoints in images. SURF is based on SIFT algorithm, a robust method proposed by Lowe in 1999 [8] to find keypoints and describe local features invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint.

Roughly speaking, SURF detects keypoints by using scale-space as an image pyramid, where the image is iteratively convolved with Gaussian kernel and repeatedly sub-sampled at different scales [9].

Those pyramid layers are subtracted in order to get the DoG (Difference of Gaussians) images where edges and blobs can be found [10]. Keypoints are local maxima/minima in a 3x3x3 neighborhood in the image over scales.

One dominant orientation is assigned to each interest point found in the image by calculating the sum of all responses from a sliding orientation window Haar filter [10]. Thus a 64 dimension vector of wavelet (Haar) responses relative to dominant orientation is extracted.

Finally, each image is composed of n keypoints and each keypoint contains local descriptor represented by a 64 dimension features vector extracted by SURF. More details can be found in [10].

3 Proposed Approach

This work proposes a feature vector containing orientation gradients information combined with color histograms to identify visual patterns of interest regions in aerial images. Such information will be available for regions classification, which allows test images to be matched only with waypoints inside the chosen region. Therefore the aim of this approach is to use regions classification as a filter which reduces the number of comparisons during image matching phase and then speed up the UAV location during navigation.

Given a set of images of an environment, colors may be an important feature to describe a regions. Considering regions are a limited area with some objects inside, the colors distribution of images from the same region tends to be similar, see Figure 1. In this sense the Hue histogram from HSV (Hue, Saturation and Value) color model is proposed to compose the region descriptor. Hue could be described as the color value by itself, it is invariant and independent from other channels in HSV model, for more details see [11].

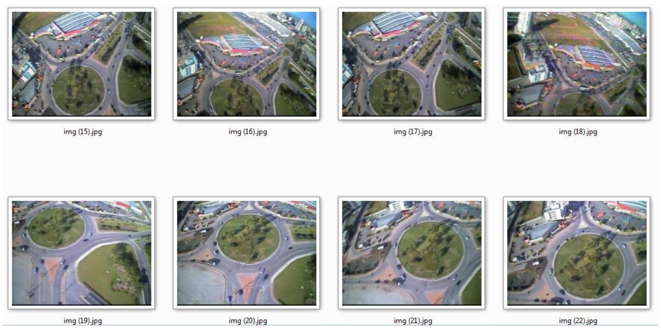

Fig. 1. Images representing two different regions. Note that only using color features it could be possible to classify them.

The Hue value is measured in degrees and it goes from 0 to 180. The histogram is composed by 30 bins, in sets of 6 degrees intervals, reducing the number of attributes in the feature vector to speed up classification time and making the histograms less variant to noise.

Many works have used color histograms to describe images [12, 13], but when color distribution in different regions are too similar the classifier could not distinguish them, since traditional histogram does not take into account spatial information.

Having this problem in mind this approach proposes in addition to the hue histogram an orientation gradient histogram, called here SURF histogram, it will compose the region descriptor in order to represent structural features of images.

SURF algorithm has successfully been used to describe local features, however the goal of region descriptor in this work is to represent a global context of the image (scene) instead of a single pixel region. In this sense a single vector of orientation features is constructed in order to represent the whole content of a scene.

The first step for constructing SURF histogram is to use SURF to find keypoints and describe local features; each keypoint has a 64 features vector of Haar wavelet responses (gradients orientation). Then the next step is to calculate the mean average of each vector value using all keypoints in the image. It will result in a single vector of 64 features per image. Since keypoints are supposed to describe common and singular objects in the images, if images are representing the same region, the average of keypoints descriptors should be similar.

Combining color and orientation features the final vector for describing regions has 94 attributes, 30 from color histogram and 64 from SURF histogram. Supervised learning algorithms can be trained with these features extracted from samples of interest regions and finally used for region classification.

4 Experimental Evaluation

Having in mind the goal of evaluating region classification step and comparing the results obtained by the selected classifiers, this section describes the details of this experiment including an analysis of results.

4.1 Training Dataset

In order to create the training dataset for experimentation, 161 aerial images were selected from the SURF recognized waypoints set in the experiment presented in [7]. In this previous work, a set of test and sample images where matched using SURF algorithm in the conventional approach where every test image is compared to all samples in the dataset. Now the goal is to validate if the images are classified in the proper region which the waypoint belongs and also estimate the improvements of time processing using this additional classification step.

In the sequence, a set of 94 attributes for each sample were extracted using SURF histogram and the Hue histogram.

For each of the 161 region examples, a feature vector was calculated and stored into the dataset. The training examples were manually labeled with one of the following classes: Region 1, Region 2, Region 3, Region 4 and Region 5, see Figure 2. The distribution of classes is: 21 images in Region 1, 31 in Region 2, 40 in Region 3, 23 in Region 4 and 47 in Region 5. The number of samples for each class was extracted from the available images of each region in the environment, see examples of sample images in Figure 3.

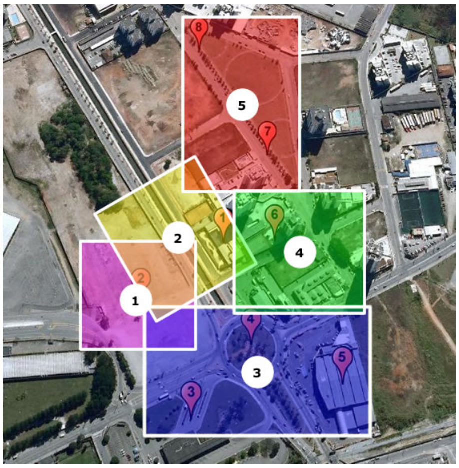

Fig. 2. Visual representation of the five regions covering all waypoints in the environment

Fig. 3. Samples: Top images contain samples of Region 3 and Bottom images samples of Region 2.

4.2 Experimental Settings

The experiments were conducted using OpenCV 1.1 for dataset building and the latest developer version of Weka software [3] [14] for classifiers evaluation.

The supervised algorithm Support Vector Machines (SVM) was tested in conjunction with the well known algorithms Multi Layer Perceptron (MLP) and K-nearest neighbors (KNN). All classifiers implementations were performed using Weka default configuration. SVM uses a polynomial function, complexity parameter separation C=1.0 and ϵ = 1.0E-12.

SVM classifier was chosen due to its generalization capability and fast classification time [15]. The experiments uses Sequential Minimal Optimization (SMO), a implementation SVM developed by John C. Platt [16], who claims that SMO is a simple and fast technique to solve the SVMs quadratic problem.

The MLP was chosen because of its ability of Neural Networks to implicitly detect complex nonlinear relationships between attributes of training data, this classifier uses backpropagation method and parameters hidden Layers = (number of attributes + number of classes), learning Rate=0.3, momentum =0.2 and trainingTime = 500, more details of these parameters can be obtained in [17].

KNN classifier was selected as a weak classifier to be compared with SVM and MLP and then evaluate the complexity of the classification problem, it uses a linear search applying Eucledian distance and parameter K=1, more details of parameters can be found in [18].

For each of the algorithms 5-fold cross validation was performed over the dataset in order to certify a more reliable estimation of the generalization error [19]. A paired-samples t-test was also conducted to compare the set of features combining color and orientation histograms.

4.3 Results and Analysis

The experiments were exploratory and conducted with the intention of evaluating the specificity (Precision), sensitivity (Recall) and efficiency (processing time x accuracy) of algorithms for region classification problem using Color and SURF based histograms. The Tables 1, 2 and 3 show the CPU training time, CPU classification time and overall accuracy of the three classifiers. Each classifier was trained and tested with three different combinations of features vector using color and orientation histograms.

Table 1. Results of SVM classifier for region classification problem using aerial images

	Color Hist.	Ori. Hist	**Train. time**	**Classif. Time**	**Accuracy**
SVM	X	-	0.1158 ms	0.0006 ms	92.5956%
SVM	-	X	0.1186 ms	0.0008 ms	92.9044 %
SVM	X	X	0.1535 ms	0.0012 ms	96.2733 %

[3] Weka is open source software which has a collection of machine learning algorithms for data mining tasks, more details in http://www.cs.waikato.ac.nz/ml/weka/

Table 2. Results of MLP classifier for region classification problem using aerial images]

	Color Hist.	Ori. Hist	**Train. time**	**Classif. Time**	**Accuracy**
MLP	X	-	12.9218 ms	0.0016 ms	92.6336%
MLP	-	X	11.6400 ms	0.0016 ms	91.6544 %
MLP	X	X	34.6597 ms	0.0106 ms	96.8944 %

Table 3. Results of KNN classifier for region classification problem using aerial images]

	Color Hist.	Ori. Hist	**Train. time**	**Classif. Time**	**Accuracy**
KNN	X	-	0.0002 ms	0.0051 ms	91.4130 %
KNN	-	X	0.0000 ms	0.0039 ms	89.7794 %
KNN	X	X	0.0000 ms	0.0218 ms	92.5466 %

In the conventional approach of image matching presented in [7], all test images are matched with all sample images in the dataset. Then the global processing time Gt is given by:

$$Gt = Tim \times Mt \times Sim \quad (1)$$

Where Tim is the number of test images, Mt is the average time of image matching process and Sim the number of sample images in the database.

Considering the region classification step proposed in this work, test images would be matched only with the samples images of the region it was classified, therefore the new global time NGt would be reduced to:

$$NGt = (\sum_{i=1}^{n} Ri \times Mt) + Tim \times Ct \quad (2)$$

Where Ri is the number of images in the region i, n is the number of regions and Ct is the classification time for one instance.

Considering Mt = 0.163 ms (according on experiments in [7]), Si=8 and Ti = 161 (number of test images in this work) the Gt = **209.944ms**. Based on the distribution of the regions and the classification time presented in Tables 1, 2 and 3, the Table 4 presents , for each classifier, the new global time NGt and the timing performance gain.

Table 4. Timing performance gain using regions classification before matching step.

	NGt (ms)	Perf. Gain ($\times$ faster)
SVM	52.353	**4.01**
MLP	53.866	3.89
KNN	55.669	3.77

The Table 5 shows the standard deviations of results for percentage correct classifications. Results marked with • are significantly different at confidence p <0.05, based on a pared t-test.

Table 5. Standard deviation results based on pared t-test. SVM and KNN classifiers are compared to MLP in the first column. The symbols ∘, • mean that the results have statistically significant improvement or degradation with confidence p <0.05

Dataset	MLP	SVM	KNN
Color H.	± 4.95	± 4.59	± 4.32
Orient. H.	± 2.56	± 3.70	± 4.52
Color+Orient.	± 2.65	± 2.71	± 3.88 •

Although MLP has performed the best classification (Table 2), its possible to see in Table 5 that the results achieved by MLP are not statistically significant better than SVM classifier. Nevertheless, the efficiency of the algorithms during the testing phase is of interest as well. Note that the processing time of testing phase of SVM is by far the best in terms of efficiency. It is justified by the fact that the time for evaluating test cases is proportional only to the final number of support vectors. Nevertheless KNN presents the best time during training and the SVM with the second best time (see Tables 3 and 1).

The classification results using the combination of color plus orientation histograms obtained the best accuracy, from the set of 161 instances, SVM classified only 6 instances incorrectly, MLP 5 and KNN 12. SVM algorithm has the best overall performance and the low numbers of instances incorrect classified only shows that all selected classifiers are very suitable for the region classification problem using aerial images.

The results confirm that the set features can discriminate the regions very accurately. Note that SURF-based histogram features could discriminate the region classes very precisely, it is possible to observe how the appearance of the orientation histogram is preserved in different samples images from the same region in Figure 4. The final features vector also including the color histogram increased significantly the classification rate that reached more than 96% of accuracy.

In the Table 6 is possible to observe the behavior of the classifiers using color and orientation histograms with respect to precision, recall, and the F-measure.

Table 6. True Positives, False positives, Precision, Recall, F-Measure and ROC curve averages for SVM, MLP and KNN algorithms

	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**ROC curve**
SVM	0.963	0.01	0.964	0.963	0.963	0.986
MLP	0.969	0.008	0.97	0.969	0.969	0.997
KNN	0.925	0.017	0.932	0.925	0.926	0.958

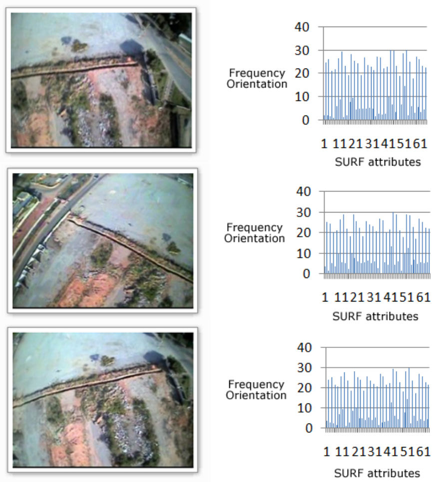

Fig. 4. SURF based histograms of aerial images from the same objects of an environment area.

5 Conclusions

The image descriptor presented in this work was able to represent global context in the set of aerial images experimented and provided excellent results with more than 95 % of accuracy. With that, waypoints can be grouped by regions reducing the processing time of recognition step during navigation. The results of experiments in this work showed a gain of performance that speeds up the global the processing up to 4 times faster. But note that some of regions evaluated here had only one or two waypoints. Therefore, this approach can be still better if considering large environments with dense grids of mapped waypoints.

In order to better evaluate the model, a natural step is the application of similar solutions at different environments and larger datasets in both urban and rural areas. It is also of interest to perform experiments using satellite images, such the ones obtained from Google Earth, for matching with UAV images.

A future research direction is the exploitation of presented image descriptors to the improvement of classification and also reduction of the number of parameters using feature selection algorithms in order decrease processing time and then to satisfy real industry needs.

References

1. SHAH, P.Z.: Pakistan says u.s. drone kills 13. New York Times (2009)
2. Valavanis, K.P.: Advances in Unmanned Aerial Vehicles: State of the Art and the Road to Autonomy. 1st edn. Springer Publishing Company, Incorporated (2007)

3. Shiguemori, E.H., Martins, M.P., Monteiro, M.V.T.: Landmarks recognition for autonomous aerial navigation by neural networks and gabor transform. In: IPAS. (2007) 64970
4. Rodrigues, R.C.B., Shiguemori, E.H., Forster, C.H.Q., Pellegrino, S.R.M.: Color and texture features for landmarks recognition on uav navigation, São José dos Campos, Simpósio Brasileiro de Sensoriamento Remoto, 14.(SBSR), Instituto Nacional de Pesquisas Espaciais (INPE) (2009) 7111–7118
5. Royer, E., Lhuillier, M., Dhome, M., Lavest, J.M.: Monocular vision for mobile robot localization and autonomous navigation. Int. J. Comput. Vision **74**(3) (2007) 237–260
6. Conte, G., Doherty, P.: An integrated uav navigation system based on aerial image matching. In: In Proceedings of the 2008 IEEE Aerospace Conference, Big Sky Montana, USA (2008)
7. Rodrigues, R.C.B., Pellegrino, S.R.M.: An experimental evaluation of algorithms for aerial image matching. In: 17th International Conference on Systems, Signals and Image Processing, Rio de Janeiro, Brazil, EdUFF (2010) 416–419
8. Brown, M., Lowe, D.: Invariant features from interest point groups. In: Proceedings of the British Machine Vision Conference 2002, BMVC 2002, Cardiff, UK, British Machine Vision Association (2002) 656–665
9. Evans, C.: Notes on the opensurf library. Technical report, University of Bristol, Bristol, UK (2009)
10. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-up robust features (surf). Computer Vision and Image Understanding **110**(3) (2008) 346 – 359 Similarity Matching in Computer Vision and Multimedia.
11. Schwarz, M.W., Cowan, W.B., Beatty, J.C.: An experimental comparison of rgb, yiq, lab, hsv, and opponent color models. ACM Trans. Graph. **6**(2) (1987) 123–158
12. Arjunan, R.V., Kumar, V.V.: Image classification in cbir systems with color histogram features. Advances in Recent Technologies in Communication and Computing, International Conference on **0** (2009) 593–595
13. Cheng, Y.C., Chen, S.Y.: Image classification using color, texture and regions. Image and Vision Computing **21**(9) (2003) 759 – 776
14. Weka Machine Learning Project: Weka. URL http://www.cs.waikato.ac.nz/˜ml/weka (2008)
15. Vapnik, V.N.: The nature of statistical learning theory. Springer-Verlag New York, Inc., New York, NY, USA (1995)
16. Platt, J.: Sequential minimal optimization: A fast algorithm for training support vector machines (1998)
17. Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice Hall PTR, Upper Saddle River, NJ, USA (1998)
18. Bremner, D., Demaine, E., Erickson, J., Iacono, J., Langerman, S., Morin, P., Toussaint, G.: Output-sensitive algorithms for computing nearest-neighbour decision boundaries. Discrete and Computational Geometry **33** (2005) 593–604 10.1007/s00454-004-1152-0.
19. Imbault, F., Lebart, K.: A stochastic optimization approach for parameter tuning of support vector machines. In: ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 4, Washington, DC, USA, IEEE Computer Society (2004) 597–600