próximo
artigo

Color and Texture Features for Landmarks Recognition on UAV Navigation

Ricardo Cezar B. Rodrigues[1]
Elcio Hideiti Shiguemori [2]
Carlos Henrique Q. Forster [1]
Sergio Roberto M. Pellegrino [1]

[1] Aeronautics Technological Institute (ITA)
P. Marechal Eduardo Gomes, 50- CEP 12.228-900, São José dos Campos, São Paulo, Brazil
{ricardo, forster, pell}@ita.br
[2] Institute for Advanced Studies (IEAv)
Rod. dos Tamoios, km 5.5 - Putim - Cep 12.228-001, São José dos Campos, São Paulo, Brazil
elcio@ieav.cta.br

Abstract. Image processing in real-time is a fundamental issue in many applications in computer vision such as remote sensing, tracking and autonomous navigation. Nowadays many unmanned aerial vehicles (UAVs) depend on Global Positioning System (GPS) and inertial systems for navigation and are controlled by a ground control station. Vision systems could improve the autonomous capacity of navigation of such vehicles. The goal of this paper is to present empirical experiments of landmark recognition using color and texture features. Once the landmarks are recognized in a class of geo-referenced images they can be used to estimate the UAV position and help the autonomous navigation. Experiments are presented here using one set of aerial geo-referenced images (train set) and another set of aerial images (test set) of the same region, collected in different time, which are not geo-referenced. Two implementations of supervised learning algorithms, namely, Neural Networks and Adaptive Boosting were used to classify the instances. Besides the accuracy, training time and testing time other three metrics, recall, precision and ROC curve, were used to evaluate the landmark recognition experiments. Even using images from complex environments with different angles, illuminations and scales, the obtained recognition rate of up to 99% indicates the adequacy of using color and texture features for landmark recognition in autonomous aerial vehicle navigation.

Keywords: pattern recognition, image processing, UAV navigation ..

1. Introduction

In most of applications, UAV navigation systems employ GPS and inertial sensors (INS). However, there are some problems with this approach, as current GPS navigation system is vulnerable to noise and atmosphere effects. In addition, the response of INS system drifts in time and will be unusable after a few seconds. A vision navigation system can be used to solve navigation related problems. There are several researches that focus the problem of solving the aerial, terrestrial or nautical vehicle autonomous navigation based on vision (KUNDUR; RAVIV, 1998) (AZINHEIRA et al., 2002). When dealing with vision-based autonomous navigation systems for unmanned aerial vehicles (UAV) several challenges have to be deal with. The vision system captures a huge amount of data that must be processed in real time so that relevant information can be extracted from frames to feed controlling and navigation systems. This information is fundamental for the safe and efficient accomplishment of the preplanned mission by the UAV.

Landmark recognition systems for UAV have strict requirements of high processing speed, limited payload for navigation systems and significant landmark variations due to factors such as humankind actions, the seasons landscape changes and sun illumination. The use of texture features in object recognition systems are widely used in many works (BHAGAVATHY; NEWSAM; MANJUNATH, 2002; LATIF-AMET; ERTUZUN; ERCIL, 2000; HSEU; BHALERAO; WILSON, 1999). Textures can define surface characteristics, in images it can be defined as a function of the spatial variation in pixel intensities (gray values) (CHEN; PAU; WANG, 2000). This work presents a landmark recognition system based on the extraction of Color and Texture features. The proposed method uses this features due to provide information about surface orientation, shape and color. This approach is under study for application in the PITER (Real-Time Image Processing) research project, carried out at the Institute for Advanced Studies (IEAv - Instituto de Estudos Avancados), and applied in autonomous UAV navigation based on images. This paper is organized as follows. Section 2 describe a few of works related to the UAV navigation problem. Section 3 presents an overview and definition of features extractors and learning algorithms. Section 4 describes the experiments and its results. Finally Section 5 presents the conclusion and future work.

2. Related Works

Most of current UAV navigation systems uses GPS signals and are controlled by a ground control station, as described in (ZHOU; ZANG, 2007). The authors in (MILLER; SHAH; HARPER, 2008) use image registration for UAV landing. The adopt a simplification of the camera model in order to estimate UAV position by measuring image geometry. Information about the terrain surrounding the runway at different scales and distances is used.

In (CONTE; DOHERTY, 2008) a similar work presents a vision system using aerial image matching for UAV position estimation. The authors propose the matching of video frames with geo-referenced image database. The system uses a visual odometer and Kalman filter fused with data from an inertial sensor. The results have encouraged the new experiments using the method proposed.

In (SHIGUEMORI; MARTINS; MONTEIRO, 2007) the authors present a landmark recognition system based on Artificial Neural Networks and Gabor Filters for UAV autonomous navigation based on images. Better results were obtained over urban area, however, INS information was considered to the aerial images pre- processing .

3. Supervised Learning Approach for landmark recognition

Supervised learning is a particular type of machine learning algorithm that allows prediction of the class a previously unknown instance based on the knowledge of the class and attribute of a training sample. This technique has been successfully employed in the solution to many significant real-world problems (MITCHELL, 2006). In a similar fashion, the authors of this paper propose the use of supervised learning for the solution to the landmark recognition problem in aerial images, the Figure 1 shows the proposed method of landmark recgnition on this work. In our approach, a classifier is trained from image regions representing landmarks in order to map a set of features extracted from the image region to the type of landmark it represents if any.

The goal of this section is to present how the landmark recognition problem can be modeled as a supervised learning problem. For this purpose, initially basic definitions are presented, followed by an overview of the feature extractors.

3.1. Basic Definitions

Regarding the supervised learning terminology, the following definitions will be considered in the context of the landmark recognition problem. A labeled instance is a pair $(\bar{x}, y)$ where $\bar{x}$ is a vector in the d-dimensional space X. The vector $\bar{x}$ represents the feature vector with $d = 130$ attributes extracted from a region within a given aerial image and y is the class label associated with $\bar{x}$ for a given instance, details on the attribute extraction phase are found in section 4.1. Therefore, a classifier is a mapping function from X to Y. The classifier is induced through a training process from an input dataset which contains a number n of labeled examples $(\bar{x}_i, y_i)$ for $1 \leq i \leq n$.

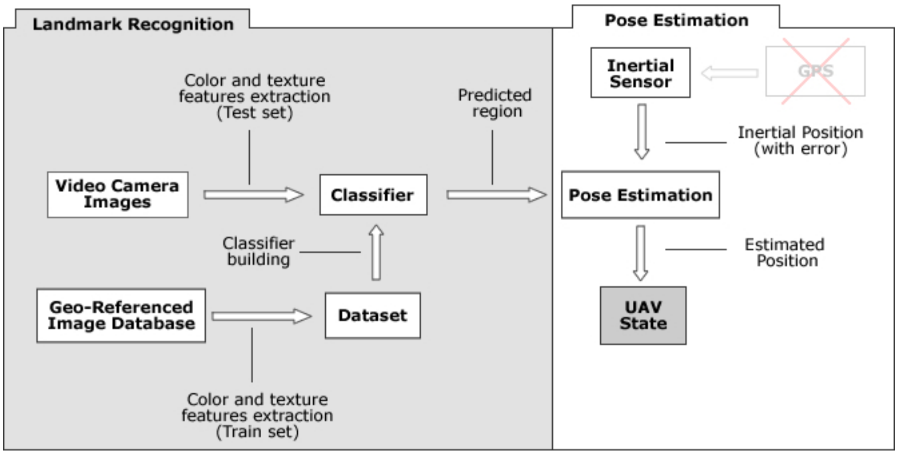

Figura 1: Landmark recognition model on UAV navigation

For the experiments, a set of five types of regions has been chosen containing the following elements: building, road, nature, landmark1 and landmark2. The use the two landmarks elements are specific regions in the image. They have been chosen because they are reference objects and then they can be used to find the location for UAV navigation. From each region extracted from the aerial images, a set of features were extracted using color and texture attributes.

In order to apply machine learning effective, we must select the most appropriate algorithms for a given problem. Two classifiers implementations are evaluated, Multilayer Perceptron (MLP) and the boosting of Decision Trees using the Adaboost (FREUND; SCHAPIRE, 1999) algorithm.

3.2. HSB and Co-occurrence Matrix

The HSB color model gives the information about hue, saturation and brightness over image. This method has advantages over other models like RGB (Reg, Green an Blue) due to invariance such as environment conditions (WANG; SUTER, 2003).

The Gray Level Co-occurrence Matrix (GLCM) one of the best known methods for texture feature extraction in images (PARTIO BOGDAN CRAMARIUC; VISA, 2002). The GLCMs describes the relation between each pixel and its neighborhood, given by the parameters, δ (distance between pixels) and θ (orientation) (JOBANPUTRA; CLAUSI, 2004). In other words, through an 8bits image matrix a new resulting matrix stores the co-occurrence values of similarity according to the parameters δ and θ. In a smooth (low contrast) image the values of neighboring pixels tend to be similar, but if the image has a high contrast checkerboard like texture, most pixels will have a neighbor with an opposite value. In this case, the co-occurrence matrix will be empty along the diagonal and largely populated near the top-left and bottom right corners. The method gives yields mesuares of, energy, entropy correlation among other (JOBANPUTRA; CLAUSI, 2004).

In order to exemplify, the Figure 2 shows one matrix m of 3×3 where each matrix position $m(i, j)$ represents a gray level pixel intensity in the range 0 to 2, wich generates a 2 x 2 result matrix. In this example, pairs of pixels have their intensity compared with angle θ equal 0 and distance of pixels δ equal 1. The resulting matrix is incremented for each pair of pixels with same intensity value. The position $(0, 0)$ of the resulting matrix, for example, contains the times that the pairs of pixels with intensity 0 were founded.

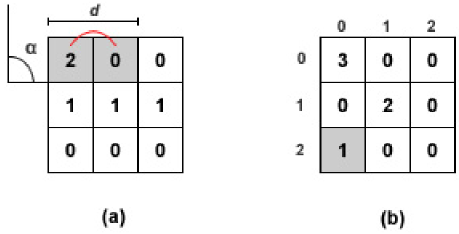

Figura 2: (a) Original matrix. (b) resulting matrix of co-occurrence

4. Experiments

With the goal of evaluating the results obtained by the selected classifiers with the feature extraction model proposed, this section describes the details about the experiments conducted for this paper along with an analysis of their results.

4.1. Dataset

In order to create the dataset for experimentation, four aerial images from an urban region were selected from the PITER repository. Two aerial images were taken in the region of Sao Jose dos Campos, Brazil. They are geo-referenced, have 1056x1056 pixels resolution and cover an area of approximately 6.35 kilometers. The other two images were taken by a helicopter in a low level flight, simulating the vision system of an unmanned aerial vehicle (UAV) in the same region, they have 720x480 pixels resolution, contain some landmarks also present in the others images, but they are not geo-referenced. The two subset of images were taken by different sensors in different times with low variation in illumination, angle and scale. The geo-referenced images in the set represents the previous knowledge about the landmarks, while the other images simulate an UAV vision.

A set of five classes has been chosen to label the regions in the images, namely, building, road , nature, landmark1 (a building) and landmark2 (a bridge). One sample of each of these classes can be visualized in Figure 4. The regions and landmarks were manually segmented using a software module from the DTCOURO software (AMORIM et al., 2006). A total of nineteen segments have been extracted from the images including examples of the previously mentioned regions classes. After the manual segmentation of classes, an algorithm implemented in the DTCOURO was used to extract windows of 40x40 pixels by scanning all the segments, the Figure 3 ilustrate the sample extracting process. Each window is an example that belongs to either one of the region class. A total of 2149 40x40 windows were created in this way.

Figura 3: Illustration of sample generation over the geo-referenced images database.

Tabela 1: Parameters used for the texture feature extraction technique used in this project. 126 texture features were extracted from each 40x40 window

	Co. Matrices
Initial Angle:	0
Final Angle:	180
Angle variation:	10
Distance variation:	1

(a) (b) (c) (d) (e)

Figura 4: Sample Example of (a) Landmark1, (b) Nature, (c) Road, (d) Building and (e) Landmark2.

The next step is the feature extraction from each 40x40 window. A set of 129 attributes for each sample were extracted , three of then use the information the average intensity of hue, saturation and brightness of the HSB color model, the rest of the attributes were extracted by the co-occurrence matrix method, the configuration are presented in the Table 1. The matrix describe image features of entropy, contrast, dissimilarity, correlation, inverse difference moment, inverse difference and angular second moment. More details about these similarity metrics can be obtained in (JOBANPUTRA; CLAUSI, 2004; AMORIM et al., 2006).

For each of the 2149 examples, a feature vector $\bar{x}$ was calculated and stored into the dataset. At the same time, all the training examples were already labeled with one of the following classes: $\{building, road, nature, landmark1, landmark2\}$, the distribution of classes is as follows: 432 $building$, 537 $road$, 486 $nature$, 613 $landmark1$ and 81 $landmark2$ examples, where the number of examples in each class is proportional to the area of each defective region in the original images.

4.2. Experimental Settings

The experiments were conducted using the latest version of the Weka software (WITTEN; FRANK, 2005). Both learning algorithm implementations, MLP (Neural Network) and Adaboost with J48 (Decision Trees) were tested. For each of the algorithms 10-fold cross-validation was performed over the dataset in order to certify a more reliable estimation of the generalization error (IMBAULT; LEBART, 2004).

4.3. Evaluation with the Supervised Learning Algorithms

The experiments are basically exploratory and were conducted with the intention of evaluating the effectiveness and efficiency of the algorithms over the landmarks recognition problem. The works in this subsection presents a feature selection and the result of supervised algorithms over the selected attributes.

Tabela 2: Testing and training time for AdaBoost and MLP.

	Testing time	Training time	Accuracy (%)
BoostJ48	0.006s	2.002s	99.023
MLP	0.002s	14.606s	97.998

4.3.1. Attribute Selection

Having in mind the improvement of training and testing time, this experiment was performed to reduce the number of features. This experiment was conducted using the Weka attributes selector BestFirst search method with Correlation-based Feature Subset Selection (CFS) for Machine Learning which evaluates the worth of a given subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them (HALL, 1998). Just 9 of the 130 attributes were selected by the CFS Subset Evaluator, where 3 are color features and 6 are texture features.

The new set of features contains the all three attributes of HSB color model, it means that the color is an important feature to discriminate different regions over the images used in that experiment. In the other 121 attributes from texture using co-occurrence matrix, just 6 were considered in the model. One possible reason for that large number of attribute excluded can be justified by the number of similarity metrics and the angle variation that they were taken. The texture extractor method uses angle variation of 10 to 10, it improves the rotation invariance, but in some cases like that many attributes can be discarded.

After the attribute selection phase, the algorithms were tested again using just the selected features. Table 2 shows the execution time for the testing and training phases as well as the respective accuracy of the two classifiers. Adaboost J48 and MLP have shown excellent and similar performance with respect to the classification task, nevertheless, the efficiency of the algorithms during the testing phase is of interest as well. Note that the testing phase of AdaBoost-J48 are by far the best in terms of efficiency. It is justified by the fact that the time for evaluating test examples is proportional to the number of base classifiers (decision trees) multiplied by the height of each decision tree.

4.3.2. Algorithms Validation

The confusion matrix is a $|Y| \times |Y|$ bi-dimensional array where the position (i, j) denotes the number of examples of class i predicted as examples of the class j. Roughly speaking, each column represents the predicted examples and each row represents the actual examples. Such matrix can be used to compare the classifiers by combining their elements into more sophisticated formulas like precision, recall and area under the ROC curve. The traditional formula for precision is:

$$P = \frac{tp}{tp + fp} \,. \tag{1}$$

where tp is the number of true positives and fp is the number of false positives. Precision is the ratio between the correctly predicted examples from a given class over the total number of actual examples of such class. On the other hand, recall is defined as the ratio between the number of correctly predicted examples from a given class and the total number of predicted examples for such class. Recall is often called sensitivity and is traditionally defined by:

$$TPR = \frac{tp}{tp + fn} \; . \tag{2}$$

where fn is the number of false negatives.

In Table 3 it is possible to observe the behavior of the algorithms with respect to precision, recall, and the area under the ROC curve. Note that all the implementation obtained relevant results.

Tabela 3: Execution results for precision, recall and area under the ROC curve.

	Roc	Recall	Precision
BoostJ48	0.999	0.995	0.980
MLP	0.998	0.990	0.972

The outstanding precision and recall values as well as the perfect area under the ROC curve demonstrate the suitability of supervised learning algorithms for the landmark recognition problem. In addition, it can be concluded that the set of features extracted from the original images boosts the effectiveness of the classifier.

5. Conclusion and Future Work

The standard UAV navigation systems generally depends on the availably GPS signal and inertial systems which system drifts in time and becomes unusable after a few seconds of use. The experiments using color and texture features and supervised learning methods presented in this work encourage the vision system as a potential solution for position estimation in the UAV navigation. As expected, the use of feature selection had presented a effective solution for the landmark recognition problem.

Two supervised algorithms were tested with the database using aerial images in different conditions of lightness, angle and scale. Note that the difference in satisfactory effectiveness between Adaboost-J48 and MLP can be neglected once both results are outstanding.

A natural step in future work is to fuse the results obtained with the landmarks recognition with a module of pose estimation based on the camera geometry methods. Another research direction is the application of similar solutions at different environments which are characterized by choosing different features.

Referências

AMORIM, W. P. et al. Desenvolvimento de um software de processamento e geracao de imagens para classificacao de couro bovino. *SIBGRAPI- Workshop of Undergraduate Works*, 2006.

AZINHEIRA, J. R. et al. Visual servo for the hovering of na outdoor robotic airship. In: *IEEE International Conference on Robotics and Automation*. [S.l.: s.n.], 2002. p. 2782–2792.

BHAGAVATHY, S.; NEWSAM, S.; MANJUNATH, B. S. Modeling object classes in aerial images using texture motifs. *icpr*, IEEE Computer Society, Los Alamitos, CA, USA, v. 02, p. 20981, 2002. ISSN 1051-4651.

CHEN, C. H.; PAU, L. F.; WANG, P. S. P. (Ed.). *Handbook of Pattern Recognition and Computer Vision*. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 2000. ISBN 9810230710.

CONTE, G.; DOHERTY, P. An integrated uav navigation system based on aerial image matching. *Aerospace Conference, 2008 IEEE*, p. 1–10, March 2008. ISSN 1095-323X.

FREUND, Y.; SCHAPIRE, R. E. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, v. 14, n. 5, p. 771–780, 1999. Disponível em: <http://www.cs.ucsd.edu/ yfreund/papers/IntroToBoosting.pdf>.

HALL, M. A. *Correlation-based Feature Selection for Machine Learning*. [S.l.], 1998.

HSEU, H.-W. R.; BHALERAO, A.; WILSON, R. G. *Image Matching Based On The Co-occurrence Matrix*. Coventry, UK, 1999. Disponível em: <citeseer.ist.psu.edu/hseu99image.html>.

IMBAULT, F.; LEBART, K. A stochastic optimization approach for parameter tuning of support vector machines. In: *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04)*. Washington, DC, USA: IEEE Computer Society, 2004. p. 597–600. ISBN 0-7695-2128-2.

JOBANPUTRA, R.; CLAUSI, D. Texture analysis using gaussian weighted grey level co-occurrence probabilities. In: *Proceedings of the Canadian Conference on Computer and Robot Vision - CRV*. [S.l.: s.n.], 2004. p. 51–57.

KUNDUR, S.; RAVIV, D. A vision-based pragmatic strategy for autonomous navigation. *Pattern Recognition*, v. 31, p. 1221–1239, 1998.

LATIF-AMET, A.; ERTUZUN, A.; ERCIL, A. An efficient method for texture defect detection: Sub-band domain co-occurrence matrices. *Image and Vision Computing*, v. 18, n. 6, p. 543–553, May 2000.

MILLER, A.; SHAH, M.; HARPER, D. Landing a uav on a runway using image registration. *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, p. 182–187, May 2008. ISSN 1050-4729.

MITCHELL, T. M. *The discipline of machine learning*. [S.l.], 2006. Disponível em: <http://www.ml.cmu.edu/CMU-ML-06-108.pdf>.

PARTIO BOGDAN CRAMARIUC, M. G. M.; VISA, A. Rock texture retrieval using gray level co-occurrence matrix. In: *In Proceedings of 5th Nordic Signal Processing Symposium*. [S.l.: s.n.], 2002. p. 795–799.

SHIGUEMORI, E. H.; MARTINS, M. P.; MONTEIRO, M. V. T. Landmarks recognition for autonomous aerial navigation by neural networks and gabor transform. In: DOUGHERTY, J. T. A. K. O. E. E. R. (Ed.). *Image Processing: Algorithms and Systems V*. [S.l.: s.n.], 2007. V, p. 6497.

WANG, H.; SUTER, D. Color image segmentation using global information and local homogeneity. *7th International Conference on Digital Image Computing: Techniques and Applications (DICTA'03)*, Sydney, p. 89–98, 2003.

WITTEN, I. H.; FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques*. second. San Francisco, CA, USA: Morgan Kaufmann, 2005.

ZHOU, G.; ZANG, D. Civil uav system for earth observation. *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, p. 5319–5322, July 2007.