Probabilistic Texture-Based Classification and Localization of 3D Objects Based on Wavelet-Features Extracted from Different Color Spaces

Marcin Grzegorzek, Alexandra Wolyniec, Frank Schmitt, and Dietrich Paulus

Institute for Computational Visualistics
University of Koblenz-Landau
Universitaetsstr. 1, 56070 Koblenz
marcin@uni-koblenz.de

Abstract. This paper presents a probabilistic approach for automatic texture-based classification and localization of 3D objects. First, the objects are described by local feature vectors extracted from color images using the wavelet transform. Later, the elements of the feature vectors are treated as random variables and modeled as density functions according to their distributions. Therefore, the objects are learned in the training phase as multimodal density functions. In the recognition phase, local feature vectors are computed from an image with an unknown object in an unknown pose. Those features are then evaluated against the trained density functions. The maximization of the evaluation values yields the classes and the poses of objects found in the image. Our experimental study on a dataset with more than 40.000 images with real heterogeneous background has shown that the classification and localization rates are dependent on the color space used for feature extraction. Therefore, the focus of this paper is to experimentally compare the RGB and Lab color spaces to each other in terms of recognition rates.

1 Introduction

One of the most fundamental problems of computer vision is the recognition of objects in digital images. The term object recognition comprehends both, classification and localization of objects. The task of object classification is to determine the classes of objects occurring in the image $\boldsymbol{f}$ from a set of predefined object classes $\Omega = \{\Omega_1, \Omega_2, \ldots, \Omega_\kappa, \ldots, \Omega_{N_\Omega}\}$. Generally, the number of objects in a scene is unknown. However, in this work we assume that exactly one object is expected in an image. In the case of object localization, the recognition system estimates the poses of objects in the image, whereas the object classes are assumed to be known. The object poses are defined relatively to each other with a 3D translation vector $\boldsymbol{t} = (t_x, t_y, t_z)^{\mathrm{T}}$ and a 3D rotation vector $\boldsymbol{\phi} = (\phi_x, \phi_y, \phi_z)^{\mathrm{T}}$ in a coordinate system with an origin placed in the image center [1].

For recognition of 3D objects in 2D images, two main approaches are known in computer vision: based on the result of object segmentation (shape-based), or by directly using the object texture (texture-based). Shape-based methods make use of geometric features such as lines or corners extracted by segmentation operations. These features as well as relations between them are used for object description [2]. However, the segmentation-based approach often suffers from errors due to loss of image details or other inaccuracies resulting from the segmentation process. Texture-based approaches avoid these disadvantages by using the image data, i. e., the pixel values, directly without a previous segmentation step. For this reason the texture-based method for object recognition has been chosen to develop the system presented in this contribution.

The object recognition problem has been intensively investigated in the past. Many approaches to object recognition, like the one presented in this paper, are founded on probability theory [3], and can be broadly characterized as either generative or discriminative according to whether or not the distribution of the image features is modeled [4]. Generative models such as principal component analysis (PCA) [5], independent component analysis (ICA) [6] or non-negative matrix factorization (NMF) [7] try to find a suitable representation of the original data [8]. In contrast, discriminative classifiers such as linear discriminant analysis (LDA) [9], support vector machines (SVM) [10], or boosting [11] aim at finding optimal decision boundaries given the training data and the corresponding labels [8]. The system presented in this paper represents the generative approaches.

Classification and localization of objects in images is a useful, and often indispensable step, for many real life computer vision applications. Algorithms for automatic computational object recognition can be applied in areas such as: face classification [12], fingerprint classification [13], handwriting recognition [14], service robotics [15], medicine [16], visual inspection [17], the automobile industry [18], etc. Although successful applications have been developed for some tasks, e. g., fingerprint classification, there are still many other areas that could potentially benefit from object recognition. The system described in this article has been tested in real application scenarios. One of these is the classification of artefacts following a visit to a museum, another is the analysis of metallography images from an ironworks.

There are further interesting approaches for object recognition. Amit et al. proposes in [19] an algorithm for multi-class shape detection in the sense of recognizing and localizing instances from multiple shape classes. In [20] a method for extracting distinctive invariant features from images that can be used to perform reliable matching between different views of an object or scene is presented. In [21] the problem of detecting a large number of different classes of objects in cluttered scenes is taken into consideration. [22] proposes a mathematical framework for constructing probabilistic hierarchical image models, designed to accommodate arbitrary contextual relationships. In order to compare different methods for object recognition, in [23] a new database specifically tailored to the task of object categorization is presented. In [24] an object recognition system is described that uses a new class of local image features. The features are invariant to image scaling, translation, and rotation, and partially invariant to illumination changes and affine or 3D projection. In [25] a multi-class object detection framework whose core component is a nearest neighbor search over object part classes is presented.

Our experimental study on a dataset with more than 40.000 real-world images has shown that the classification and localization rates are dependent on the color space which is used for feature extraction. Therefore, in this paper we experimentally compare the RGB and Lab color spaces to each other. The paper is structured as follows. Section 2 presents the training phase of the system, Section 3 deals with the classification and localization, Section 4 describes and discusses the results, and finally, Section 5 concludes the paper.

2 Statistical Training

For training, we extract feature vectors using both, the texture and the color of the objects. For the object modeling we use either RGB or Lab color spaces and compute six-dimensional local feature vectors. The main advantage of the local feature vectors is that a local disturbance affects only the features in a small region around it. In contrast to this, a global feature vector can totally change, if only one pixel in the image varies. The system determines a set of local feature vectors for all training images using the discrete wavelet transform. The Johnston wavelet and its corresponding scaling function are used for this purpose.

Some feature vectors $\boldsymbol{c}_{\kappa,m}$ describe the object Ω_κ, others belong to the background. In a real world environment, it cannot be assumed that the background in the recognition phase is a-priori known. Therefore, for the statistical object modeling, only feature vectors describing the object should be considered. Since the object takes usually only a part of the image, a tightly enclosing object area O_κ is learned for each object class Ω_κ in the training phase.

Finally, for all object classes Ω_κ considered in a particular recognition task, statistical object models $\mathcal{M}_\kappa$ are learned in the training phase. The models are regarded as continuous functions defined on the pose parameter domain $\mathcal{M}_\kappa(\boldsymbol{\phi}, \boldsymbol{t})$. This means that the object models $\mathcal{M}_\kappa$ contain the object area O_κ and thereby the set of object feature vectors $\boldsymbol{c}_{\kappa,m}$, the density functions for the object feature vectors $p(\boldsymbol{c}_m|\boldsymbol{\mu}_{\kappa,m}, \boldsymbol{\sigma}_{\kappa,m})$, and the density value for the background features $p(\boldsymbol{c}_m) = p_b$ for all pose parameters $(\boldsymbol{\phi}, \boldsymbol{t})$ in the continuous sense.

3 Statistical Recognition

Since for all object classes Ω_κ regarded in a particular recognition task corresponding object models $\mathcal{M}_\kappa$ have already been learned in the training phase, the system is able to classify and localize objects in images taken from a real world environment. First, a test image is taken, preprocessed, and feature vectors in it are computed. Second, the system starts the recognition algorithm integrated into it.

The object classification and localization for single-object scenes is solved based on the so-called maximum likelihood estimation. First, local feature vectors are extracted from the preprocessed test image with the wavelet transform, whereas both RGB and Lab color space are separately taken into consideration. Second, the object area O_κ is determined for all class κ and pose $(\boldsymbol{\phi}_h, \boldsymbol{t}_h)$ hypotheses using the learned object models $\mathcal{M}_\kappa$. Only feature vectors $\boldsymbol{c}_m$ inside the object area are taken into account for evaluation of each hypothesis. Finally, the density values $p(\boldsymbol{c}_m|\boldsymbol{\mu}_{\kappa,m}, \boldsymbol{\sigma}_{\kappa,m})$ for the object feature vectors $\boldsymbol{c}_m$, which are greater than the background density p_b, are multiplied by each other. The result of this multiplication is normalized by a quality measure called geometric criterion and maximized over all class and pose hypotheses in order to find the optimal class $\widehat{\kappa}$ and pose $(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{t}})$ for the test image.

4 Experiments and Results

For experiments, an image database for 3D object recognition in a real world environment (3D-REAL-ENV) was generated. This database consists of ten objects. The training images of these objects were taken on dark background under two different illuminations from 1680 viewpoints. Thus, there are altogether 33600 training scenes. For testing of the system, three types of test scenes, namely 2880 test images with homogeneous background, 2880 test images with less heterogeneous background, and 2880 test images with more heterogeneous background from 288 different viewpoints were acquired (see Figure 1). Illumination in the test images is different from the illumination conditions in the training scenes. The test viewpoints, in general, are also different from the training points of view. Additionally, more than 200 different real heterogeneous backgrounds were used for acquiring test images. Due to all these properties, the task of object classification and localization is very difficult for the 3D-REAL-ENV image database.

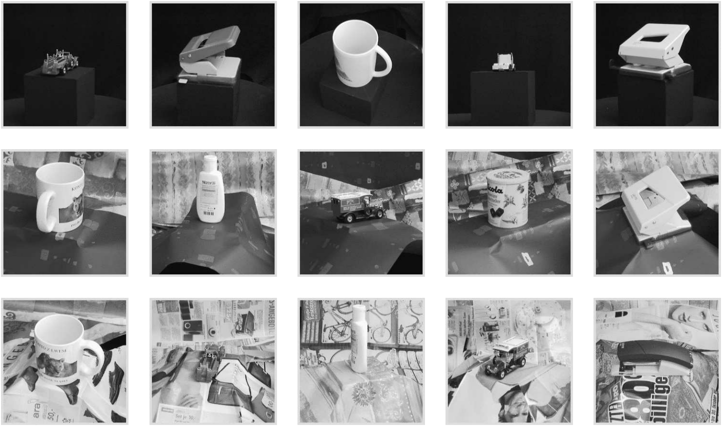

Fig. 1. Example single-object test images from the 3D-REAL-ENV database [26]. First row: test images with homogeneous background. Second row: test images with less heterogeneous background. Third row: test images with more heterogeneous background.

In the preprocessing step the images have been transformed into either RGB or Lab color spaces. Further steps of the training and the recognition phase remained same for both color spaces. As can be seen in Table 1, the classification and localization rates are dependent on the color space used.

5 Conclusions

The system proposed in this paper is featured by high performance in varying real-world environments. However, during our evaluation process we have observed a strong dependency between the recognition results and the color space in which images are coded.

References

1. Grzegorzek, M., Reinhold, M., Niemann, H.: Feature extraction with wavelet transformation for statistical object recognition. In Kurzynski, M., Puchala, E., Wozniak, M., Zolnierek, A., eds.: 4th International Conference on Computer Recognition Systems, Rydzyna, Poland, Springer-Verlag, Berlin, Heidelberg (May 2005) 161–168
2. Latecki, L.J., Lakaemper, R., Wolter, D.: Optimal partial shape similarity. Image and Vision Computing Journal **23** (2005) 227–236

Recognition Rates for 3D-REAL-ENV Image Database							
Distance of Training Views [°]	Type of Object Modeling	Classification Rate [%]			Localization Rate [%]		
		Hom. Back.	Less Het. Back.	More Het. Back.	Hom. Back.	Less Het. Back.	More Het. Back.
4.5	LAB	100	67.0	90.2	99.1	80.9	69.0
	RGB	100	88.0	82.3	98.5	77.8	73.6
9.0	LAB	100	66.7	90.7	98.7	80.0	67.2
	RGB	100	88.3	81.2	98.2	76.4	72.1
13.5	LAB	100	65.3	87.8	96.9	78.6	65.4
	RGB	99.6	82.7	80.3	94.9	68.4	66.6
18.0	LAB	100	63.3	83.9	96.6	71.4	54.5
	RGB	97.3	80.6	68.6	94.3	64.9	60.7
22.5	LAB	99.9	57.5	77.7	94.5	60.7	38.6
	RGB	94.7	74.8	59.2	89.4	52.2	46.2
27.0	LAB	99.1	43.6	62.2	83.8	49.9	32.8
	RGB	93.8	53.6	50.2	78.3	35.8	35.6

Table 1. Classification and localization rates obtained for 3D-REAL-ENV image database with gray level and color modeling. The distance of training views varies from 4.5° to 27° in 5 steps. For experiments, 2880 test images with homogeneous, 2880 test images with less heterogeneous, and 2880 images with more heterogeneous background were used.

3. Schiele, B., Crowley, J.L.: Recognition without correspondence using multidimensional receptive field histograms. International Journal of Computer Vision **36**(1) (January 2000) 31–50
4. Ulusoy, I., Bishop, C.M.: Generative versus discriminative methods for object recognition. In: International Conference on Computer Visions and Pattern Recognition (Volume 2), San Diego, USA, IEEE Computer Society (June 2005) 258–264
5. Jolliffe, I.T.: Principal Component Analysis. Springer (2002)
6. Hyvarinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons (2001)
7. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature **401** (1999) 788–791
8. Roth, P.M., Winter, M.: Survey of appearance-based methods for object recognition. Technical Report ICG-TR-01/08, Inst. for Computer Graphics and Vision, Graz University of Technology, Austria (2008)
9. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. John Wiley & Sons (2000)
10. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer (1995)

11. Freund, Y., Shapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer System Sciences **55** (1997) 119–139
12. Gross, R., Matthews, I., Baker, S.: Appearance-based face recognition and light-fields. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(4) (April 2004) 449–465
13. Park, C.H., Park, H.: Fingerprint classification using fast fourier transform and nonlinear discriminant analysis. Pattern Recognition **38**(4) (April 2005) 495–503
14. Heutte, L., Nosary, A., Paquet, T.: A multiple agent architecture for handwritten text recognition. Pattern Recognition **37**(4) (April 2004) 665–674
15. Zobel, M., Denzler, J., Heigl, B., Nöth, E., Paulus, D., Schmidt, J., Stemmer, G.: Mobsy: Integration of vision and dialogue in service robots. Machine Vision and Applications **14**(1) (April 2003) 26–34
16. Li, C.H., Yuen, P.C.: Tongue image matching using color content. Pattern Recognition **35**(2) (February 2002) 407–419
17. Ngan, H.Y., Pang, G.K., Yung, S., Ng, M.K.: Wavelet based methods on patterned fabric defect detection. Pattern Recognition **38**(4) (April 2005) 559–576
18. Gausemeier, J., Grafe, M., Matysczok, C., Radkowski, R., Krebs, J., Oelschlaeger, H.: Eine mobile augmented reality versuchsplattform zur untersuchung und evaluation von fahrzeugergonomien. In Schulze, T., Horton, G., Preim, B., Schlechtweg, S., eds.: Simulation und Visualisierung, Magdeburg, Germany, SCS Publishing House e.V. (March 2005) 185–194
19. Amit, Y., Geman, D., Fan, X.: A coarse-to-fine strategy for multi-class shape detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(12) (December 2004) 1606–1621
20. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2) (November 2004) 91–110
21. Torralba, A., Murphy, K.P., Freeman, W.T.: Sharing visual features for multiclass and multiview object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(5) (May 2007) 854–869
22. Jin, Y., Geman, S.: Context and hierarchy in a probabilistic image model. In: IEEE Conference on Computer Vision and Pattern Recognition, New York, USA (June 2006) 2145–2152
23. Leibe, B., Schiele, B.: Analyzing contour and appearance based methods for object categorization. In: IEEE Conference on Computer Vision and Pattern Recognition, Madison, USA (June 2003)
24. Lowe, D.G.: Object recognition from local scale-invariant fearures. In: 7. International Conference on Computer Vision (ICCV), Corfu, Greece (September 1999) 1150–1157
25. Mahamud, S., Hebert, M.: The optimal distance measure for object detection. In: IEEE Conf. on Computer Vision and Pattern Recognition, Madison, USA (June 2003)
26. Reinhold, M., Grzegorzek, M., Denzler, J., Niemann, H.: Appearance-based recognition of 3-d objects by cluttered background and occlusions. Pattern Recognition **38**(5) (May 2005) 739–753