DISCRIMINANT SPLITTING FOR FEATURE EXTRACTION APPLICABLE TO VIEW-INDEPENDENT OBJECT RECOGNITION

Ioannis Marras, Nikos Nikolaidis, Ioannis Pitas

Informatics and Telematics Institute, Centre for Research and Technology Hellas, Greece
Department of Informatics, Aristotle University of Thessaloniki, Greece
E-mail(s): {imarras, nikolaid, pitas}@aiia.csd.auth.gr

ABSTRACT

This paper presents a novel object recognition technique that is based on discriminant image splitting for feature extraction. Spatially homogeneous and discriminant regions for each object class are generated. The classical image splitting technique is used in order to determine those regions. Thus, each object class is characterized by a unique region pattern which consist of homogeneous and discriminant 2-D regions. The mean intensities of these regions are used as features. The proposed method has been tested in view independent object recognition at the COIL-100 database. The obtained results demonstrate that the method can achieve a very satisfactory recognition rate.

1. INTRODUCTION

Object recognition constitutes one of the most active areas in computer vision with applications in many fields such as robot navigation, multimedia understanding and semantic description and surveillance. The majority of the proposed algorithms are based on finding local image patches that are robust or invariant to image scaling, partial occlusion or affine transformations. For this purpose, viewpoint-invariant interest point detectors [1, 2] in combination with appropriate patch descriptors [3, 4] are used. In [2], a rate-efficient codec that compresses scalable single view tree structures that describe the hierarchy of SIFT histograms used for recognition is proposed, while in [5] a multiple-view SIFT feature selection algorithm is proposed for object recognition. Several methods have been developed to deal with the problem of model-based object recognition by solving the correspondence problem through tree search. Many global invariants that are used to describe the model have been introduced in the literature. These include the affine invariant moments, as well as other techniques such as cross-weighted moments [6], affine invariant spectral signatures [7] and the trace transform [8]. However, all of the above mentioned methods appear to be either computationally expensive or difficult to implement.

Recent approaches include Multi-Scale Autoconvolution (MSA) [9] and Spatial Multi-scale Affine invariants (SMA) [10]. In [11], a new computational model scheme that merges color and shape invariant information for object recognition is proposed. Furthermore, shape invariants are computed based on moments, Fourier transform coefficients, edge curvature and arc length [12, 13]. In appearance-based object recognition, the majority of work is based on shape information and incorporates matching sets of shape image features (like edges, corners and lines) between a query and a target image [13].

Motivated by the use of graphs with nodes placed at discriminant points for face recognition [14], we propose an object recognition technique that segments the object images to discriminant regions. The main idea is the creation of a set of regions that is discriminative for each class in the sense that a subset of these discriminant and homogeneous regions will provide adequate information in order to distinguish a certain object class from another one. The entire set is necessary in order to distinguish this class from the rest of the other classes. The region segmentation is based on a variant of the classical image splitting technique. The features that this method uses are the mean intensities of the produced regions.

This paper is organized as follows: the discriminant splitting approach that is used during the training phase in order to extract the characteristic features for each class is presented in Section 2, for both single-view and view-independent object recognition. The actual recognition (classification) procedure and the experimental evaluation of the method on the COIL-100 multi-view object database are presented in Sections 3 and 4, respectively. Conclusions follow.

2. FEATURE EXTRACTION USING A DISCRIMINANT SPLITTING TECHNIQUE

2.1 Single View Object Recognition

In the simplest case, we assume that there exist n object classes and each class contains l different, equally sized, centered images of the object taken approximately from the same view. Thus, the dataset D is divided into n sets, each containing the samples of an object, $D = \bigcup_{i=1}^{n} \mathscr{U}_i$. The main goal is to find homogeneous regions that are discriminant between the classes. In this way, for each class, a unique regions pattern, i.e. a set of regions, is created. This procedure, that is based on a splitting approach, will be described bellow.

Let two classes a, b each containing l samples (images) of the corresponding object, in sets $\mathscr{U}_a$, $\mathscr{U}_b$. If each image is of dimensions $h \times w$, these l images can be considered as a stack of slices (volume) with dimensions $l \times h \times w$, as illustrated in Figure 1. Thus for our purpose, a certain region B can be considered as being a parallelepiped volume comprising of the parts of every image in the class that fall within the region. We assume that an image I is divided into R regions. For a region B defined as above and for a class a we define its discriminant power, with respect to class b, using the Fisher's discriminant ratio [15] that is:

$$F_{a,b}(B) = \frac{(\mu_a(B) - \mu_b(B))^2}{\sigma_a(B)^2 + \sigma_b(B)^2}, \tag{1}$$

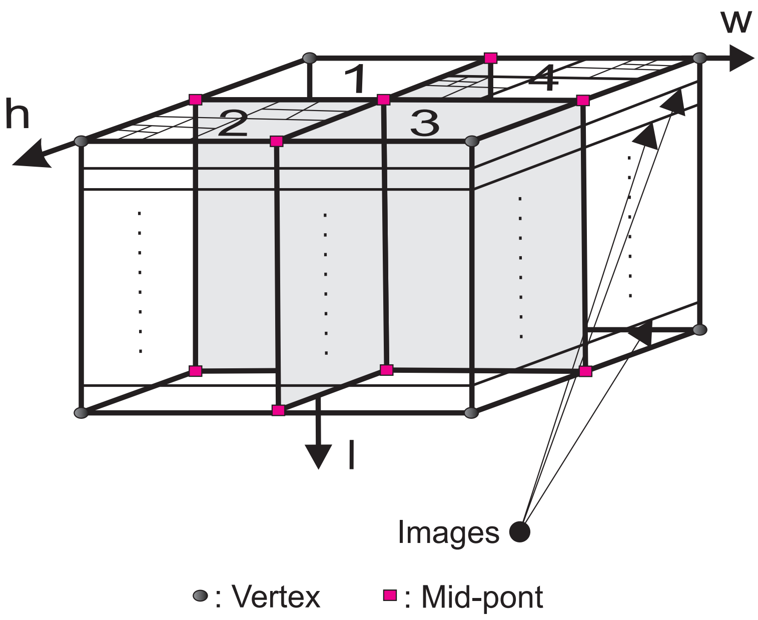

Figure 1: Class representation as a stack of images.

where $\mu_a(B)$ and $\sigma_a(B)$ are the mean intensity and variance for the region B of all the images that belong to class a, while $\mu_b(B)$ and $\sigma_b(B)$ are the mean intensity and variance for the same region of the images that belong to class b. A region B_1 is more discriminant than a region B_2, for a particular object pair a, b, when $F_{a,b}(B_1) > F_{a,b}(B_2)$. As mentioned above, except from the discriminant power of a region the method exploits also its homogeneity. Any volume homogeneity check method can be used. We have chosen the one based on the intensity range $|I_{max} - I_{min}|$, where I_{max}, I_{min} are the maximum and minimum intensity values of a region. If the range is smaller than a certain threshold, i.e.:

$$|I_{max} - I_{min}| \leq T_s, \tag{2}$$

then the region is regarded to be homogeneous, where the threshold T_s denotes the Otsu threshold [16] calculated for the current region. As in the case of region discriminant power calculation, the homogeneity of a region is judged based on the pixels intensity values of the parts of *all* the class's training images that fall within the region's boundaries, i.e. on all pixels of the corresponding volume.

In order for the discriminant and homogeneous regions to be determined for each class a, the classical splitting approach is applied to the l images of this class. For each object class the stack of images is recursively split into four quadrants or regions (Figure 1), until 2D discriminant and non-homogeneous regions are encountered. The splitting is performed by bisecting the rectangular regions (in the entire image stack) in the vertical and horizontal directions.

The splitting procedure for the stack of slices of a class a proceeds as follows: For a region B under consideration we evaluate the discriminant ratio in (1) for all pairs of classes, i.e. we evaluate $F_{i,j}(B)$, $i, j = 1 \dots n$, $i \neq j$. We then find the largest ratio $F_{max}(B) = F_{i^*,j^*}(B) = \max_{i,j}(F_{i,j}(B))$. If this maximum involves the region B in class a, i.e. if $i^* = a$ or $j^* = a$ this means that B is most discriminant in the task of distinguishing a from the other classes. In this case, region B is split for class a. Otherwise, the homogeneity of B in a is examined using (2) and the region is split if it is inhomogeneous, until homogeneous regions are reached, otherwise it is not split. In summary, if a region is very discriminant for a class it is being split, whereas if it is not discriminant enough, it is split if it is inhomogeneous.

The above procedure is performed for each object class separately. In the end of the training procedure, a region pattern for each class is created. Four object images along with the corresponding region patterns are shown in Figure 2.

Finally, each training image I_k within a class a is characterized by the vector $\underline{\mu}_{ak}$ that contains the mean intensity values for each of the regions $r_{a,j}$, $j = 1 \dots n_a$, n_a being the number of regions in the pattern of class a. The rationale behind the splitting procedure outlined above for the creation of the region pattern for each class is that since each region is finally described by its mean value, large regions are described in a very coarse way, since they are represented by a single value, whereas an area split into many small regions is represented in a more refined and detailed way, as every such small region is described by its own mean value. Thus, the algorithm splits regions that are discriminant for a certain class into smaller regions, in order to describe these regions with finer detail, which is important for classification due to their discriminant power. The fact that the method also splits non-discriminant regions that are inhomogeneous helps the fine tuning of the region placement in the testing (classification) procedure, as will be explained in the next section.

2.2 View-Independent Object Recognition

For view independent object recognition, a number of l images are available for each of a number of different views of an object. Let n_{view} denote the number of different views for each of the n objects. In this case the algorithm utilizes $n_{Total} = n \cdot n_{view}$ classes. If an object image is classified (see next section) to one of the n_{view} classes (essentially sub-classes) of object class a, then the image is assigned to this object class.

3. IMAGE CLASSIFICATION

The algorithm's testing (classification) procedure is as follows. The n_a discriminant regions r_{aj} of class a are selected upon an image I depicting a certain object from an arbitrary view. In order to solve possible alignment problems, the pattern (set) of regions for each class is not considered as a rigid pattern, but the regions boundaries are translated locally by small amounts until they fall on as much as possible homogeneous regions. For a class a, the intensity means $\mu_{Ir_{aj}}$ of every region r_{aj} of class a are computed in I, providing the image I means vector $\underline{\mu}_{Ia}$. The image means vector $\underline{\mu}_{Ia}$ is then compared with the (pre-computed) means vectors $\underline{\mu}_{ak}$ for all training images I_k ($k = 1 \ldots l$) of object class a, resulting in distances $d_{Ia_k} = \|\underline{\mu}_{Ia} - \underline{\mu}_{ak}\|$ for every training image k that belongs to class a. Thus, l distances are computed for each class. This is repeated for all n_{Total} classes resulting into n_{Total} distances. The object is classified to the class α^*, that contains the training image κ^* whose means vector is closest to the test image mean vector,

$$(\alpha^*, \kappa^*) = \arg\min_{\alpha,\kappa} d_{Ia_k}. \tag{3}$$

The small region translations mentioned above are also involved in this search for the matching training image and class.

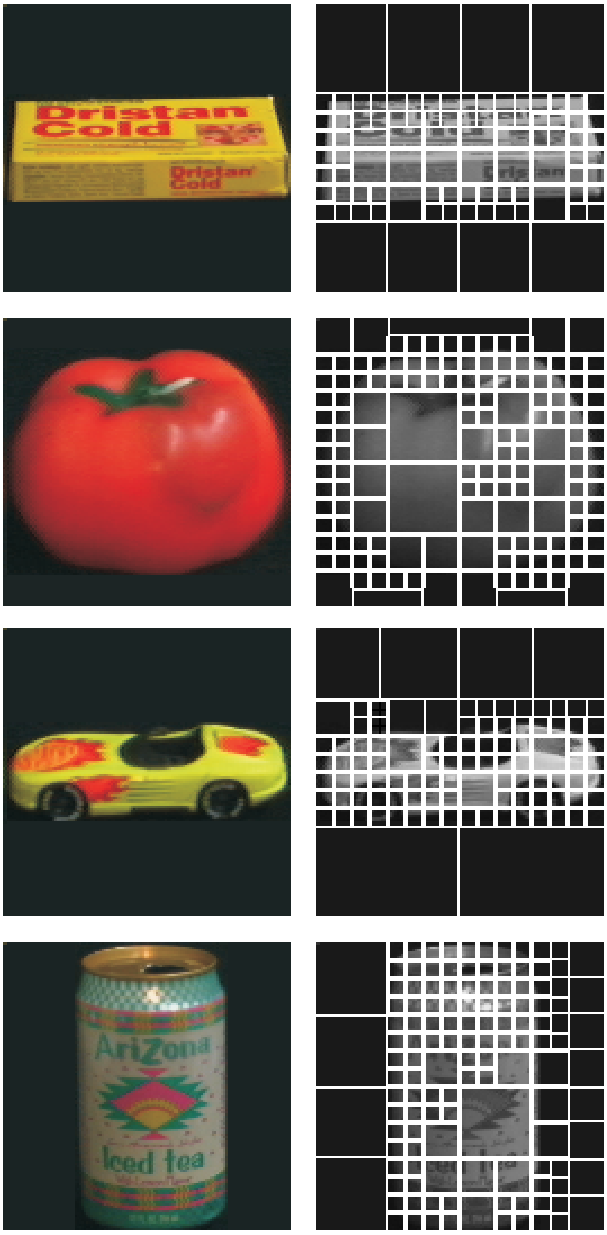

Figure 2: Four object images from the COIL-100 database along with the corresponding region patterns.

4. EXPERIMENTAL RESULTS

The proposed method for view independent object recognition was evaluated using the COIL-100 multi-view object database [17] consisting of 72 colour views (128x128 pixels) for each one of its 100 objects (Figures 2,3) viewed from 0 to 360 degrees in 5 degree increments. This dataset has been widely used in object recognition experiments. We split the COIL-100 dataset into training and testing sets so that 48 views of each object are used for the training procedure, while the remaining 24 views are used for the testing procedure. For each one of the 100 objects, during the training procedure, 16 sub-classes are created, each one containing 3 nearby views. Totally, 100*16=1600 classes have been used for our experiments.

The proposed algorithm was found to be able to recognize the objects with very satisfactory accuracy compared to other state of the art methods, like [5], applied on the same dataset. The recognition score of the proposed method was 98.5% while in [5] the recognition score for the same setup is 99.0%. In terms of computational complexity, 3.2 seconds are required for the proposed algorithm in order to categorize an object image on an Intel Pentium 4 (3.01 GHz) processor PC with 1.5GB of RAM.

5. DISCUSSION AND CONCLUSIONS

A novel method for view independent object recognition performing discriminant region splitting is proposed in this paper. Spatially non-overlapping regions compose a unique region pattern for each class. The mean intensity values of each of the regions in this pattern are used to characterize this class and classify an unknown image. Results show that the proposed technique is able to recognize the objects with very

Figure 3: Several views of an object from the COIL-100 database.

good accuracy. Furthermore, the proposed approach could be used in other image classification tasks.

6. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Community Seventh Framework Programme (FP7/2007-2013) under grant agreement no 211471 (i3DPost).

REFERENCES

[1] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, “A comparison of affine region detectors,” in *IJCV*, vol. 65(12), 2005, pp. 43–72.

[2] D. M. Chen, S. S. Tsai, V. Chandrasekhar, G. Takacs, J. Singh, and B. Girod, “Tree histogram coding for mobile image matching,” in *In Data Compression Conference*, Snowbird, UT, 16-18 March 2009, pp. 143–152.

[3] D. Lowe, “Object recognition from local scale-invariant features,” in *ICCV*, 2005.

[4] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, “Surf: Speeded up robust features,” in *Computer Vision and Image Understanding (CVIU)*, vol. 110(3), 2008, pp. 346–359.

[5] C. Christoudias, R. Urtasun, and T. Darrell, “Unsupervised feature selection via distributed coding for multiview object recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, AK, 23-28 June 2008 2008, pp. 1–8.

[6] Z. Yang and F. Cohen, “Cross-weighted moments and affine invariants for image registration and matching,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 804–814, August 1999.

[7] J. Ben-Arie and Z. Wang, “Pictorial recognition of objects employing affine invariance in the frequency domain,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 6, pp. 604–618, June 1998.

[8] M. Petrou and A. Kadyrov, “Affine invariant features from the trace transform,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 30–44, January 2004.

[9] E. Rahtu, M. Salo, and J. Heikkila, “Affine invariant pattern recognition using multiscale autoconvolution,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 908–918, June 2005.

[10] ——, “A new efficient method for producing global affine invariants,” in *Proceedings of the International Conference on Image Analysis and Processing (ICIAP)*, Cagliari, Italy, 2005, pp. 407–414.

[11] A. Diplaros, T. Gevers, and I. Patras, “Combining color and shape information for illumination-viewpoint invariant object recognition,” *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 1–11, January 2006.

[12] I. Weiss, “Geometric invariants and object recognition,” *International Journal of Computer Vision*, vol. 10, no. 3, pp. 207–231, 1993.

[13] T. Reis, *Recognizing Planar Objects using Invariant Image Features*. Springer- Verlag, Berlin, 1993.

[14] S. Zafeiriou, A. Tefas, and I. Pitas, “Learning discriminant person-specific facial models using expandable graphs,” *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 1, pp. 55–68, 2007.

[15] K. Fukunaga, *Statistical Pattern Recognition*. Academic Press, San Diego, CA, 1990.

[16] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, pp. 62–66, 1979.

[17] S. A. Nene, S. K. Nayar, and H. Murase, “Columbia object image library (coil-100),” Technical report, Colmubia University, Tech. Rep., 1996.