

Figure 5: Sample of the Object Database.

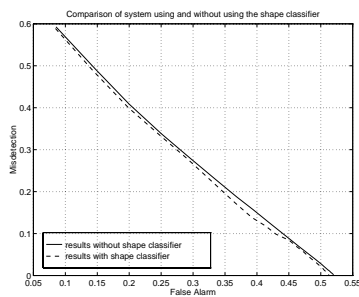


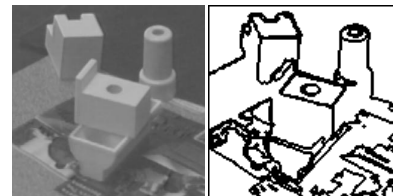
Figure 6: Misdetection and false alarms with 10% of occlusion, with and without using shape groups.

parison of appearance and geometric model based recognition. In *Lecture Notes in Computer Science (1144): Object Representation in Computer Vision II*. Springer-Verlag, 1996.

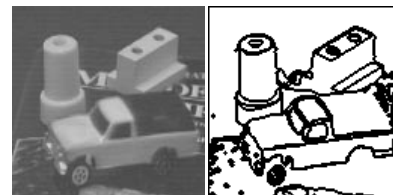
- [7] H. Murase and S. K. Nayar. Visual Learning and Recognition of 3-D Objects from Appearance. *International Journal of Computer Vision*, 14:5–24, January 1995.
- [8] C. A. Rothwell. *Object Recognition through Invariant Indexing*. OUP, 1995.
- [9] S. C. Zhu and A. Yuille. Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(9):884–900, September 1996.
- [10] A. Zisserman, D. Forsyth, J. Mundy, C. Rothwell, J. Liu, and N. Pillow. 3D object recognition using invariance. *AI Journal*, (78):239–288, 1995.



Identity	Score	Pose
SpCar	0.61	302.77°
Truck	0.19	4.29°
Ambulance	0.10	11.52°



Identity	Score	Pose
CCube	1.36	41.12°
Stamp	0.89	0.64°
HoleCube	0.89	14.47°
Sink	0.68	152.94°



Identity	Score	Pose
Stamp	1.91	93.66°
TwoHole	0.96	181.63°
Truck	0.04	93.67°

(c)

Figure 7: (a) Cluttered scenes. (b) MDL segmentations. (c) Recognition results.

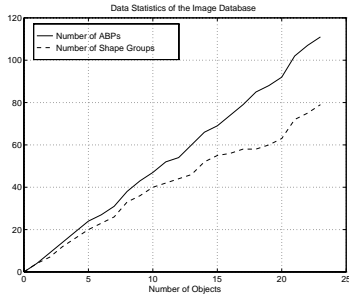


Figure 3: Number of ABPs and shape groups as the number of objects in the database increase.

7.2 Discriminant Power

Experiments were conducted by randomly selecting an appearance of a random ABP from the database, and randomly occluding 10% of its bounding box with the appearance of another ABP, also randomly selected. The projection of the occluded region is assigned the closest manifold in the system. If the distance between the projection of the region and the manifold is larger than a threshold, T , it is said that the region is *mis-detected*. If this distance is less than the threshold T , but the assigned manifold is not the one corresponding to the true identity of the region being used, it is said that this is a *false alarm*. Otherwise, it is said that the region is correctly identified. Figure 6 shows plots of misdetection versus false alarms for 10,000 experiments, as the threshold T varies from 0.015 to 0.15, with and without the use of shape groups. It is seen that the system performs better when the hierarchical structure of eigenspaces is used.

7.3 Recognition in cluttered scenes

Figure 7 shows three examples of cluttered images with occlusion, their respective MDL segmentations, and the recognition results. All the objects scene were correctly identified and localize, in spite of the occlusion and segmentation errors.

8 Conclusions

A discriminatory power index for ABPs and ABRs based on probabilistic models was proposed. This index provides quantitative measures of the goodness of the hypotheses generated by the recognition system. Furthermore, it can be used to organize databases of objects to share similar ABPs and ABRs reducing the effective size of the databases without lossing performance.

References

[1] J.-L. Chen and G. C. Stockman. Determine Pose of 3-D Objects with Curved Surfaces. *IEEE Transactions on*

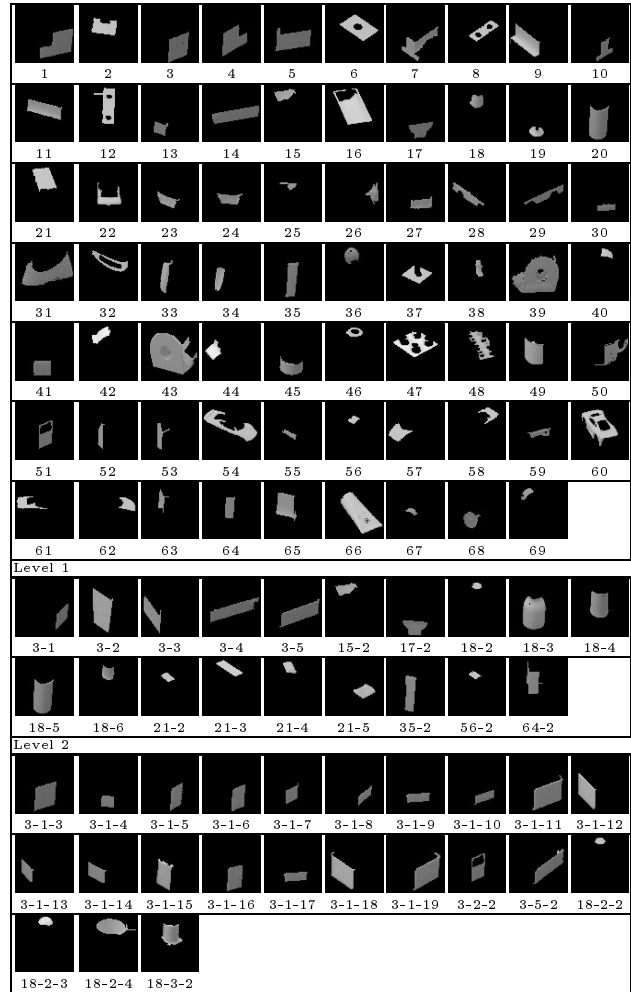


Figure 4: ABP hierarchical structure.

Pattern Analysis and Machine Intelligence, PAMI-18:57-62, January 1996.

- [2] C.-Y. Huang, O. I. Camps, and T. Kanungo. Object Recognition Using Appearance-Based Parts and Relations. In *IEEE Conference on Computer Vision and Pattern Recognition*, page to appear, June 1997.
- [3] T. Kanungo, B. Dom, W. Niblack, and D. Steele. A Fast Algorithm for MDL-based Multi-band Image Segmentation. In *Proc. IEEE Computer Vision and Pattern Recognition*, pages 609-616, Seattle, Washington, June 1994.
- [4] J. Krumm. Eigenfeatures for planar pose measurement of partially occluded objects. In *Proc. IEEE Computer Vision and Pattern Recognition*, pages 55-60, San Francisco, California, June 1996.
- [5] A. Lenardis and H. Bischof. Dealing with occlusions in the eigenspace approach. In *Proc. IEEE Computer Vision and Pattern Recognition*, pages 453-458, San Francisco, California, June 1996.
- [6] J. Mundy, A. Liu, N. Pillow, A. Zisserman, S. Abdallah, S. Utcke, S. Nayar, and C. Rothwell. An experimental com-

to their discriminant power in the eigenspace spanned by all the ABPs of “HoleCube.”

In this hierarchy, FEs are organized in a tree structure, as shown in Figure 2, by grouping together manifolds with similar DPI. Let \mathcal{M} be a set of manifolds of the same type:

$$\mathcal{M} = \{M_1, \dots, M_n\}$$

The set \mathcal{M} is partitioned into subsets $\mathcal{M}_1, \dots, \mathcal{M}_C$ such that

$$\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \dots \cup \mathcal{M}_C$$

$$\mathcal{M}_i \cap \mathcal{M}_j = \emptyset \quad \forall i \neq j$$

$$DPI(M_i|M_j) \leq \delta_k \text{ and } DPI(M_j|M_i) \leq \delta_k$$

$$\forall M_i, M_j \in \mathcal{M}_k, \quad k = 1, \dots, C.$$

Similar appearances from similar manifolds are clustered together and are aggregated into a single manifold passing through the means of these clusters.

At the top level of the hierarchy, all the FEs are clustered as described above and span the *Universal* eigenspace; then, at each node in the tree the groups are divided into subgroups, each of them consisting of more similar (with lower discriminant power indices) features. In turn, each of these subgroups defines a new eigensubspace where the highest order coordinates provide more discriminatory power between the FEs spanning it.

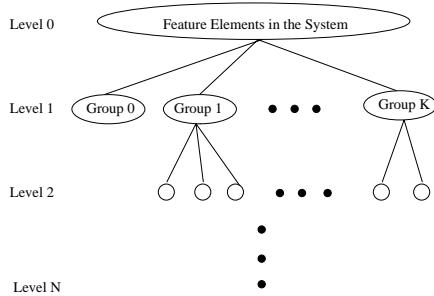


Figure 2: Hierarchical organization of ABPs.

Thus, in this hierarchical structure 1) the leaves of the tree correspond to *parts* rather than objects, thus allowing objects to share parts; 2) the similarity measure is based on probabilistic models learned from segmentations of *real* data; and 3) the discriminant power between manifolds at each level of the structure increases. Finally, the structure can be incrementally updated whenever a new FE needs to be learned, by comparing it with the existing groups, and only retraining the most similar group.

6 Combining Hypotheses

The hierarchy of ABPs and ABRs allows the system to not only hypothesize the identity of the segmented

regions, but it also provides pose hypotheses. Furthermore, each hypothesis has associated with it a probability measuring its reliability. These probabilities are combined using the following Bayesian framework.

Consider the database of objects

$$\mathcal{DB} = \{\mathcal{D}_1, \dots, \mathcal{D}_N\},$$

sharing the set of model feature elements

$$\mathcal{M} = \{M_1, \dots, M_C\},$$

and consider the set of image feature elements

$$\mathcal{S} = \{s_1, \dots, s_n\}.$$

Let $P(\mathcal{D})$ be the probability that model \mathcal{D} is observed, and let $P(h(s) = M)$ be the probability that s is an observation of feature manifold M . Assuming that all objects are equally likely, we have

$$P(\mathcal{D}) = \frac{1}{N}$$

and

$$P(\mathcal{D}|M) = \frac{P(M|\mathcal{D})P(\mathcal{D})}{\sum_k P(M|\mathcal{D}_k)P(\mathcal{D}_k)}$$

Assuming that the probability $P(h(s) = M)$ is exponential with respect to the distance between the manifold M and the point s , $d(s, M)$, we have

$$P(M|s) = \frac{1}{\sum_i e^{-kd(s, M_i)}} e^{-kd(s, M)}$$

Then, the feature hypotheses can be combined into overall Bayesian scores for each model [1]:

$$R(D) = \sum_i \sum_j P(D|M_i)P(M_i|s_j)$$

7 Experiments

7.1 Database Organization

Figure 3 shows plots of the number of ABPs and groups of ABPs as the number of objects in the database increase. It is seen that while both numbers grow approximately linearly with the number of objects, the slope for the numbers of groups is less steep. It is expected that as more objects are added to the database, the growth of the number of groups will be even slower. The current database consists of 110 ABPs grouped into 69 “shape groups” between three levels as shown in Figure 4, and 130 ABRs. Thus, the use of hierarchical grouping resulted in a 37% reduction of the ABP database. These ABPs and ABRs were generated from image sequences of 24 objects of which 16 are shown in Figure 5.

the input data the identity of the closest manifold in the database (identity mapping), and the pose of the closest point on this manifold (localization mapping): [2, 7]:

$$h_{ABP}(p) = \arg \min_{P \in \mathcal{ABP}} d(p, P), \quad (2)$$

$$l_{ABP}(p) = \arg \min_{a_P \in h_{ABP}(p)} d(p, a_P), \forall p \in \mathcal{S}_1 \quad (3)$$

$$h_{ABR}(r) = \arg \min_{R \in \mathcal{ABR}} d(r, R), \quad (4)$$

$$l_{ABR}(r) = \arg \min_{a_R \in h_{ABR}(r)} d(r, a_R), \forall r \in \mathcal{S}_2 \quad (5)$$

However, since it is possible that more than one manifold is close to the projection of the given input, making the choice of the nearest manifold somewhat arbitrary. A more reasonable approach is to consider more than one possible assignment, such that all projections within a given threshold are considered as hypotheses. Such thresholds can be found by minimizing the probabilities of false alarm and misdetection using an experimental procedure as the one presented in [2]. However, the thresholds found in this way are global, in the sense that they are the same for every manifold and every appearance on the manifolds. This observation leads us to believe that a recognition system should use information about the ABPs and ABRs discriminant power, where the discriminant power is measured in terms of how far the considered manifold is to other manifolds of the same type. In this way, the thresholds can be varied along manifolds, depending on their discriminant power, and assignments can be ranked according to their probability of being correct. Figure 1 illustrates this concept. In Figure 1(a),

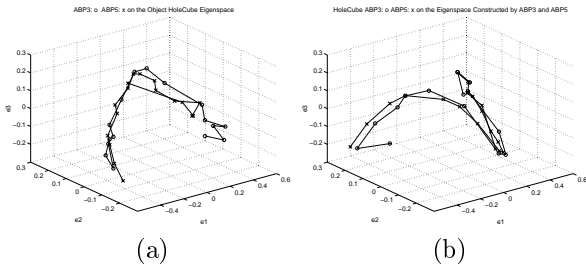


Figure 1: Similar ABPs have low relative discriminant power.

the manifolds ABP3 and ABP5 of the first object in Table 1, “HoleCube”, corresponding to the two rectangular faces shown (c) and (e) column in Table 1, are shown in the “HoleCube” eigenspace¹. Any image region whose projection in this eigenspace is close to one of the manifolds, will be as close to the other one. To quantify this type of situation we define a “discriminant power” index as follows.

We model the ABPs and ABRs as being the result of a stochastic process, where each manifold can be expressed as a nominal manifold plus noise. The results

¹The manifolds are shown as three dimensional for visualization purposes.

described below apply to both ABPs and ABRs, and hence we talk of manifolds and appearances without referring to their particular type².

Let M_k denote an instance of the k^{th} manifold of a given type and a_{kj} denote its j^{th} appearance. The probability of misclassifying appearances of M_k as appearances of M_i is given by

$$P(M_i|M_k) = \frac{\sum_j \sum_m P(a_{im}|a_{kj})P(a_{kj})}{P(M_k)}$$

This probability should be large if the manifolds M_i and M_k are close to each other for most of the appearances of M_k . Assuming that all appearances of a given type are equally likely we have:

$$P(a_{kj}) = \frac{1}{\text{Total \# appearances}}$$

$$P(M_k) = \frac{\# \text{ appearances of } M_k}{\text{Total \# appearances}}$$

Assuming that the probability $P(a_{im}|a_{kj})$ can be modeled as exponentially decreasing with the distance between the appearances a_{im} and a_{kj} :

$$P(M_i|M_k) = \frac{1}{\# \text{ frames in } M_k} \sum_j \sum_k A_{kj} e^{-kd(a_{im}, a_{kj})}$$

where

$$A_{im} = \frac{1}{\sum_j e^{-kd(a_{ij}, a_{im})}}$$

Finally, the discriminant power index, DPI, is defined as

$$DPI(M_i|M_k) = 1 - P(M_i|M_k)$$

























In this way, the discriminant power index of M_k from M_i is large whenever the probability of misclassifying M_k as M_i is low.

5 Organizing Large Databases

A good object representation should be able to scale up to be used with a large database of objects. The discriminant power index introduced above can be used to group together similar ABPs and ABRs (feature elements, or FEs) into a hierarchical organization. Since each object can potentially have many ABPs and ABRs, this is a very important property. Furthermore, by spanning new sub-eigenspaces using similar parts, it is possible to actually increase the discriminant power between these parts. For example, when the ABPs ABP3 and ABP5 of “HoleCube” shown in Figure 1(a) are used to span a new from eigensubspace as shown in Figure 1(b), their relative discriminant powers raise from 0.125 and 0.059 to 0.353 and 0.375, respectively, an increase factor of 6 with respect

²A manifold and an appearance can be either of ABP or ABR type.

Table 2: ABRs Database Sample. The ABRs of each object are represented by one of their appearances.

Object	ABRs Representatives								
	r1	r2	r3	r4	r5	r6	r7	r8	r9
									
									
									

3 Identity and pose hypotheses

Given an image and its MDL segmentation, ABP and ABR hypotheses can be generated by projecting each segmented region and pair of regions into the eigenspaces obtained during training, and finding points on manifolds near to these projections.

Let ABP and ABR be the sets of the union of the ABPs and ABRs manifolds, respectively, for *all* the objects in a given database. Then, the object database can be represented by a set of relational descriptions:

$$DB = \{\mathcal{D}_1, \dots, \mathcal{D}_N\}$$

where

$$\mathcal{D}_m = \{\mathcal{R}_1^m, \mathcal{R}_2^m\}$$

is the relational description for object m ,

$$\mathcal{R}_1^m = \{P_1, \dots, P_{m_1}\} \subseteq ABP$$

is the set of all the ABP manifolds, P_i , $i = 1, \dots, m_1$, of model m and

$$\mathcal{R}_2^m = \{R_1, \dots, R_{m_2}\} \subseteq ABR$$

is the set of all the ABR manifolds, R_i , $i = 1, \dots, m_2$, of model m . The set \mathcal{R}_1^m is a *unary* relation of parts, while the set \mathcal{R}_2^m is a *binary* relation between parts – i.e. $R_i = (P_{i_1}, P_{i_2})$, $i = 1, \dots, m_2$, $1 \leq i_1, i_2 \leq m_1$.

An MDL segmentation of an image can be described using a similar relational representation:

$$\mathcal{D}_i = \{\mathcal{S}_1, \mathcal{S}_2\}$$

where

$$\mathcal{S}_1 = \{p_1, \dots, p_{n_1}\}$$

is the set of the projections of the parts or image regions into the ABP eigenspace, and where

$$\mathcal{S}_2 = \{r_1, \dots, r_{n_2}\}$$

is the set of the projections of pairs of adjacent regions into the ABR eigenspace.

An image is an *observation* of a subset of the models. Then, the *recognition* problem is to find two unknown correspondence mappings

$$\begin{aligned} h_{ABP} &: \mathcal{S}_1 \rightarrow ABP \\ h_{ABR} &: \mathcal{S}_2 \rightarrow ABR \end{aligned}$$

associating ABPs and ABRs to image regions and pairs of regions, respectively, and the *localization* problem is to find two unknown correspondence mappings

$$\begin{aligned} l_{ABP} &: \mathcal{S}_1 \rightarrow \mathcal{R}_1^m \\ l_{ABR} &: \mathcal{S}_2 \rightarrow \mathcal{R}_2^m \end{aligned}$$

associating *appearances* of ABPs and ABRs to image regions and pairs of image regions, respectively.

The mappings h_{ABP} and h_{ABR} represent a set of ABP and ABR *identity* hypotheses while the mappings l_{ABP} and l_{ABR} represent a set of ABP and ABR *pose* hypotheses. These hypotheses constrain each other.

Let r be the projection of a pair of adjacent image regions with projections p_1 and p_2 and let $h_{ABR}(r) = R \in \mathcal{R}_2^m$ and $l_{ABR}(r) = a_R \in R$ be its ABR identity and pose hypotheses, respectively. If the ABP hypotheses for p_1 and p_2 , $h_{ABP}(p_1) = P_1$ and $h_{ABP}(p_2) = P_2$, are such that $P_1, P_2 \in \mathcal{R}_1^m$ and $R = (P_1, P_2)$ we say that the ABR hypothesis for r is *compatible* or verifies the ABP hypotheses for p_1 and p_2 . Furthermore, if the ABP identity hypotheses are compatible with the ABR hypothesis *and* the ABP pose hypotheses for p_1 and p_2 , $l_{ABP}(p_1) = a_{P_1} \in P_1$ and $l_{ABP}(p_2) = a_{P_2} \in P_2$, are such that a_{P_1} and a_{P_2} correspond to the same pose, we say that the ABR pose hypothesis for r is compatible or verifies the ABP pose hypotheses for p_1 and p_2 .

Finally, let $d(p, q)$ represent a distance metric between two points p and q in a given eigenspace and let the distance between a point p and a manifold M be defined as the distance between the point p and the closest point to p on the manifold, $d(p, M) = \min_{q \in M} d(p, q)$. Then, the distances between the projections of the image regions and pair of regions and the corresponding manifolds and appearances $d(p, h_{ABP}(p))$, $d(r, h_{ABR}(r))$, $d(r, l_{ABR}(r))$, and $d(p, l_{ABP}(p))$ are quantitative measures of the goodness of these hypotheses, with the smaller the distance, the better the match.

4 Discriminatory Power Index

Until now, the most common choice for the correspondence mappings has been to assign to the projection of

the question of whether it can be used to model and successfully recognize large number of objects needs to be addressed.

In this paper we propose to use a discriminatory power index for ABPs and ABRs to address this problem. The proposed index is defined based on noise models of the data and it is a quantitative measure of the dis-similarity between the appearances of the features. This index is used to organize large number of ABPs from the object database in a hierarchical structure of eigenspaces where different objects share similar parts, thus significantly reducing storage and recognition time requirements. Furthermore, the proposed organization defines “specialized” eigenspaces spanned by similar shapes, resulting in better recognition performance in the presence of data uncertainty.

The paper is organized as follows. In the next section, we briefly summarize the definition of ABPs and ABRs. Then, we formalize the recognition and localization problem and introduce the discriminatory power to quantify ABP and ABR similarity. Next, a hierarchical organization of the representation based on this index is described. It is shown that it not only improves discrimination among similar features but also allows to group similar shapes. Finally a Bayesian framework combining part and relations hypotheses is discussed and experimental results are shown.

2 ABPs and ABRs

Appearance-based parts and relations were defined in [2] in terms of closed regions and the union of these regions, respectively. The regions are segmented using the MDL principle, by modeling each region as a polynomial of unknown degree in the image coordinates with additive zero mean Gaussian noise of unknown covariance, whose boundaries are encoded using a chain code representation. Although this model works best for constant albedo regions, it can be easily adapted to textured regions by using a set of texture filters like the ones used by Zhu and Yuille in [9].



















Let $\Omega = \{\omega_j\}$ denote the image segmentation into regions $\{\omega_j\}$ and let Y represent the image data. Assuming that the image comes from a stochastic process that can be characterized as a polynomial gray scale surface plus Gaussian noise described by a vector of parameters β , then the MDL objective function to optimize is given by:

$$L(Y, \Omega, \beta) = L(\Omega) + L(\beta|\Omega) + L(Y|\Omega, \beta). \quad (1)$$

where the first term is the length of encoding the region boundaries, the second term is the length of encoding the parameters and the last term is the length of encoding the residuals.

The appearance of the parts and relations for different sensor and illumination sources is obtained from collection of images under varying conditions. Two parts segmented from two images of the same object obtained with similar sensor and illumination configurations, are said to be appearances of the *same part*

Table 1: ABPs Database Sample. The ABPs of each object are represented by one of their appearances.

Object	ABPs Representatives				
	a	b	c	d	e
					
					
					

if they are judged to have similar polynomial approximations in similar image locations.

Let Y_i be an $n_i \times 1$ column vector with the gray scale pixel values in part ω_i . Let d be the order of the polynomial used to fit the parts, and $m = (d + 1)(d + 2)/2$ be the number of polynomial coefficients. Let Φ_i be an $n_i \times m$ matrix of m basis functions for each of the n_i pixels – i.e. products of powers of pixel coordinates. Finally, let Θ_i be an $m \times 1$ column vector with the *optimal* regression coefficients for ω_i . Using these definitions, we have [3]

$$Y_i = \Phi_i \Theta_i + \Psi_i$$

where Ψ_i is a vector of zero mean Gaussian noise with covariance $\sigma^2 I$, and Θ_i is estimated by minimizing the fitting error:

$$\epsilon_i = \|Y_i - \Phi_i \Theta_i\|$$

Then, two parts ω_1 and ω_2 obtained from two images of the same object with different, but similar, sensor and illumination configurations, are considered appearances of the same part ω if

$$\epsilon_{1,2} = \frac{1}{n_1} \|Y_1 - \Phi_1 \Theta_2\| + \frac{1}{n_2} \|Y_2 - \Phi_2 \Theta_1\| \leq T_\epsilon$$

and

$$\Delta_{1,2} = \|\mu_1 - \mu_2\| \leq T_\Delta$$

where μ_1 and μ_2 are the centroids of the parts and T_ϵ and T_Δ are given thresholds. Note that these thresholds can be set according to the estimated noise covariance matrix $\sigma^2 I$ and the known difference in sensor locations. Furthermore, this criteria can handle both, over and under, segmentation problems by assigning more than one part in one frame to a part in the other frame.

Finally, ABPs and ABRs are compactly represented by parametric manifolds by using the Karhunen-Loeve compression method in the two eigenspaces spanned by the parts and the relations. Tables 1 and 2 show representative appearances for the ABPs and ABRs of three objects, respectively.

Since the ABP and ABR representation is learned from segmented images, it is robust to segmentation problems. Furthermore, since it is based on regions rather than on global properties, it is robust to occlusion.

Hierarchical Organization of Appearance-based Parts and Relations for Object Recognition *

Octavia I. Camps^{1,2}, Chien-Yuan Huang¹

¹Department of Electrical Engineering

²Department of Computer Science and Engineering

The Pennsylvania State University

University Park, PA, 16802

{camps,huang}@whale.ee.psu.edu

Tapas Kanungo

Center for Automation Research

University of Maryland College Park, MD 20742-3275

kanungo@cfar.umd.edu

Abstract

In [2] a new object representation using *appearance-based parts* and *relations* to recognize 3D objects from 2D images, in the presence of occlusion and background clutter, was introduced. Appearance-based parts and relations are defined in terms of closed regions and the union of these regions, respectively. The regions are segmented using the MDL principle, and their appearance is obtained from collection of images and compactly represented by parametric manifolds in the eigenspaces spanned by the parts and the relations. In this paper we introduce the discriminatory power of the proposed features and describe how to use it to organize large databases of objects.

KEY WORDS: object representation, object recognition, appearance-based methods, shape classification.

1 Introduction

The appearance of a 3D object in a 2D image depends on its shape, its reflectance properties, its pose in the scene, and the sensor and illumination characteristics. Murase and Nayar [7] have proposed an appearance-based representation to recognize 3D objects from 2D images. In this approach, the representation is learned from sequences of images, and thus can be used to learn generic objects viewed and illuminated from different orientations. Object translation and scaling, on the other hand, are taken care by normalizing the image size using the bounding box of the object. However, this representation is highly sensitive to clutter and partial occlusion, since the learning and recognition processes require the isolation of the object of interest.

Mundy et al [6] presented an experimental comparison between Murase and Nayar's method (SLAM) and two geometric model-based recognition methods described in [8] (Lewis) and [10] (Morse). This study

concluded that: 1) appearance models have the advantage that they do not require formal models to describe objects while geometric approaches rely on formal models to derive pose invariant properties; and 2) the major drawbacks of SLAM are that it is very sensitive to segmentation, in particular occlusion, that it does not lend itself well to object categorization, and that incidental variations in appearance such as texture or surface albedo must be modeled as separate objects.

Recently, there has been a significant effort devoted to try to overcome the problems caused by occlusion and background clutter to appearance-based representations [5, 4, 2]. For example, in [5] a robust method to compute the coefficients to project an image into the parametric eigenspace was presented. This method extracts the coefficients by considering subsets of image points with a hypothesis-and-test paradigm and selecting the best hypothesis by using the MDL principle. As a result, the coefficients are robust to image outliers and in particular to occlusion. However, a major problem with this technique is that it cannot handle object translation and scaling. This is because this approach works only if the dimensions of the training and testing images are equal, and the pixel locations of the object do not change at recognition time. Unfortunately, occlusion has a direct impact on the object bounding box preventing the use of image size normalization in this case. In [4] Krumm propose to handle occlusion by using small neighborhoods as features. Although this technique can handle object translation, it also suffers from scaling – i.e. it assumes that the object size in the image is the same at recognition and training time.

In [2] a new representation using appearance-based *parts* (ABPs) and *relations* (ABRs) was introduced. This representation uses local rather than global appearances, with the local regions being automatically determined using the MDL principle. This approach significantly improved the representation robustness to segmentation problems and occlusion without compromising scaling, since the regions are normalized in size using their bounding boxes. However, since the representation decomposes objects into small parts,

*This work was supported in part by NSF grants IRI9309100 and IRI9712598.