Appearance based Generic Object Recognition

Gaurav Jain, Mohit Agarwal
Department of Electrical Engineering
Indian Institute of Technology, Delhi
Hauz Khas, New Delhi 110016

Ragini Choudhury *
INRIA Rhone-Alpes
ZIRST - 655 avenue de l'Europe
Montbonnot, St. Martin
Ragini.Choudhury@inrialpes.fr

Santanu Chaudhury
Department of Electrical Engineering
Indian Institute of Technology, Delhi
Hauz Khas, New Delhi 110016
santanuc@cse.iitd.ernet.in

Abstract

In this paper we have presented a scheme for categorization of objects based upon appearance information. The appearance information has been encoded using principal component based representations. A new scheme has been suggested for constructing appearance hierarchy through combination of the appearance space of individual objects. A categorization algorithm has been presented which exploits the features of the appearance hierarchy. Experimentations have produced encouraging results.

1 Introduction

Generic object recognition strategies endeavor to recognize objects based upon coarse, protypical representations taking into account possible variabilities of the object appearance. In this paper we have presented a generic object recognition scheme using appearance based representation.

Appearance based techniques are being widely used for object recognition because of their inherent ability to exploit image based information. Appearance based object recognition methods make recognition systems easily trainable from visual data. These systems typically operate by comparing a two-dimensional, image-like representation of object appearance against prototypes stored in the memory, and finding the closest match. A class of appearance based methods make use of a lower dimensional subspace of the higher dimensional representation memory for the purpose of comparison. Murase et. al. [5] propose an approach based upon principal component analysis. In this approach, appearance information is stored in the form of uncorrelated components and the object recognition is done by finding the nearest neighbor of the projections of unknown images using the Euclidean Distance norm. Pentland. et al. [8] have used view-based and modular eigenspaces for face recognition. In [2] similar approach has been used for parametric modeling of shape. However, the problem of object categorization using appearance representation has remained largely unaddressed.

For generic object recognition a class of methods has been proposed which exploits functional role of the 3D objects. The categorization scheme reasons about form and function of 3D objects [7]. In [3], an approach has been suggested for categorization based upon similarities with multiple class prototypes. Another scheme that has been proposed is to build up visual classes to classify the objects on the basis of field histograms [6]. However, none of these methods have the provision for building up the category hierarchy using purely image based information. In [1], rejectors are defined to eliminate a large section of the candidate classes. These rejectors can be combined to form composite rejectors so that finally the recognition algorithm is applied to a smaller set. Although it reduces the number of candidates, it is still far from solving the pattern recognition problem which exists for the remaining data set. In this paper, we try to address this aspect by classifying hierarchically so that at each stage we have a smaller number of candidate classes with an increasing degree of specificity. We have proposed a method of efficiently building a category hierarchy for generic object recognition using PCA based representation of appearance information. The recognition scheme uses a decision criterion based upon reconstruction error for the purpose of categorization. The final outcome by our method is the classification of a test image. Experimentations have established utility of the method.

The remaining paper is organized as follows. In section 2, we present a mechanism for building up the appearance class based hierarchy for the representation of these objects. Section 3 describes the classification algorithm based on the appearance hierarchy that we have proposed. The results of the performance of the classification mechanism have been presented in section 4. Section 5 concludes the paper.

*The author acknowledges support from Dept. of EE, IIT, Delhi

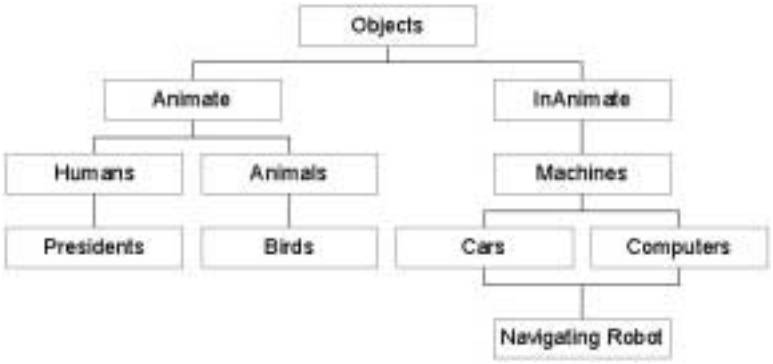

Figure 1: A sample hierarchy

2 Representation of Objects and Object Collection

In this section we describe the scheme proposed for collection and representation of different views of a generic object. This is facilitated by the generation of a compact eigen space based representation of the different views of these objects [5]. This, in fact provides an optimal representation of all the salient features of the object.

The appearance spaces of the individual objects are considered as elemental units in our collection. We propose to group these units into semantically meaningful categories. These categories, organized in a hierarchy can support query processing, particularly for QBE (query by example) scenario where we often need to categories the given example for retrieving similar objects. Hierarchy helps in finding the nearest generic category in the absence of an exact match. In this section, we describe the mechanism adopted for building up the hierarchical appearance space.

2.1 Hierarchical Combination of Appearance Space

We consider representation of a generic object in terms of eigen vectors of its appearance space. We propose to combine eigen vectors of semantically related objects for constructing appearance space of the categories. In this section we show that construction of the appearance space using multiple view of an object is equivalent to that of using eigen vectors of the appearance space of each that object.

A particular view of an object is taken as an $m \times n$ image. This is written in the form of a $mn \times 1$ vector called *image vector*. Using multiple views of the generic object, we get a set of image vectors. *Elementary sets* consist of image vectors corresponding to different views of the object.

Let

$$C_1 = \{f_1, f_2, \cdots, f_{n_1}\}$$
$$C_2 = \{g_1, g_2, \cdots, g_{n_2}\}$$

be two elementary sets, such that $C_1 \cap C_2 = \phi$. The $f_i's$ and $g_i's$ are image vectors corresponding to each view of the object. Appearance space of an elementary set can be obtained by Principal Component Analysis (PCA) [5]. Consider one such set C_1. The corresponding set of orthonormal basis vectors are denoted by

$$W_{C_1} =< e_1, e_2, \cdots, e_{nm} >$$

A *derived appearance class* is the one which is obtained by doing a PCA on two or more appearance classes.

2.1.1 Reconstruction Error

Since we are characterizing classes of objects by appearance classes, we need to address the problem of reconstruction error incurred when the reconstruction is with respect to these appearance classes. Define

$$\widehat{A} = [e_1\ e_2\ \cdots\ e_{nm}]$$

The projected vectors are

$$\widehat{f_i} = \widehat{A^t} f_i, \quad i = 1, 2...n_1$$

Now given $\widehat{A}$ and $\widehat{f_i}, i = 1, 2, ...n_1$, we can reconstruct $f_i, i = 1, ...n$ as $\widehat{A^t}^{-1} \widehat{f_i}$ (here $\widehat{A}$ is invertible as all the eigenvalues are used and these are linearly independent).

If all the eigen vectors are used then the reconstruction is exact and the reconstruction error is zero. In order to use the minimum information, we use lesser number of eigen vectors. We define

$$\widetilde{A} = [e_1\ e_2\ \cdots\ e_k]$$

Figure 2: Model Images used to build up the representation space

and

$$\widetilde{W_{C_1}} =< e_1, e_2, \cdots, e_k >$$

$\widetilde{W_{C_1}}$ is a subspace of W_{C_1}. Then the projected vectors are the $k \times 1$ vectors

$$\widetilde{f}_i = \widetilde{A^t} f_i, i = 1, 2, \cdots, nm$$

Now if we wish to reconstruct f_i from $\widetilde{f}_i, i = 1, 2, \cdots nm$ and $\widetilde{A}$, we get an approximation, say f_i^* to the image vector. Then

$$f_i^* = \widetilde{A}^{*t} \widetilde{f}_i$$

where $\widetilde{A}^{*t}$ is the generalized inverse of $\widetilde{A}^t$. The reconstruction error is given by $RE = ||f_i - f_i^*||$. By property of PCA, the reconstruction error is minimum if $\widetilde{W_{C_1}}$ consists of eigen vectors corresponding to the largest eigen values. The choice of this is to be empirically decided. For our case we have used eigen vectors which contain 90% of the information. This has been discussed in greater detail in section 3.

If $f_i \in W_{C_1}$, the reconstruction error RE is bounded by $RE \leq \Sigma_{i=k+1}^{nm} \lambda_i$ where $\lambda_i s$ are the smaller eigenvalues of AA^t. But an image vector f in general, is a vector in R^{mn}, we will characterize the reconstruction of one such vector in general, with respect to W_{C_1}, say, when $f \in W_{C_1}$. The best reconstruction, with the least reconstruction error, is given by the *bestapproximation* of f in the space W_{C_1}. The next theorem defines the best approximation [4].

Theorem 2.1 *Let W be a subspace of the space R^{mn} and let f be a vector in R^{mn}, then*

1. *The vector $f^* \in W$ is the best approximation to f by vectors in W iff $f - f^*$ is orthogonal to every vector in W.*
2. *If a best approximation to f exists, it is unique.*
3. *If W is finite-dimensional and $\mathcal{B} = \{e_1, e_2, \cdots, e_{mn}\}$ is any orthonormal basis for W, then the vector*

$$f^* = \Sigma_k \frac{(f|e_k)}{||k||^2} e_k$$

is the best (unique) approximation to f by vectors in W.

In our case, $W = \mathcal{B} = W_{C_1}$, and hence a best approximation can be found. The reconstruction error $RE = ||f - f^*||$. By the definition of *best approximation* $||f - f^*|| \leq ||f - h_i||$ for all $h_i \in W_{C_1}$.

2.1.2 Class hierarchy

The class hierarchy can be established as follows :
We consider two sets C_1 and C_2 as defined above. We construct their appearance classes W_{C_1} and W_{C_2} respectively. Here

$$W_{C_1} = \{e_1, e_2, \cdots, e_{nm}\}$$

$$W_{C_2} = \{e_1', e_2', \cdots, e_{nm}'\}$$

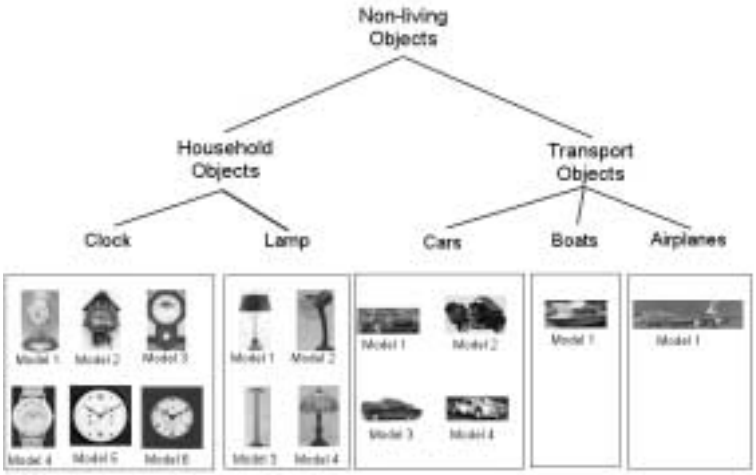

Figure 3: A sample hierarchy

Consider $D = W_{C_1} \cup W_{C_2}$. We do a PCA on this and obtain the corresponding appearance class W_D. The eigenvectors $e_1, e_2, \cdots, e_{nm}$ and $e'_1, e'_2, \cdots, e'_{nm}$ corresponding to the appearance classes W_{C_1} and W_{C_2} are able to span the appearance class W_D. Again take the set $C = C_1 \cup C_2$ and obtain its appearance class W_C.

The following theorem establishes the basis for building up the class hierarchy.

Theorem 2.2 *Let $C_1 = \{f_1, f_2, \cdots f_{n_1}\}$ and $C_2 = \{g_1, g_2, \cdots, g_{n_2}\}$ are two sets of image vectors such that $C_1 \cap C_2 = \phi$. Let W_{C_1} and the W_{C_2} be the corresponding appearance classes, computed by doing a PCA. Now we construct the appearance class of $C = C_1 \cup C_2$ say W_C and of $D = W_{C_1} \cup W_{C_2}$ say W_D. Let $f \in R^{mn}$ be any image vector. Then the reconstruction error when f is reconstructed with respect to the appearance classes W_C and W_D are the same.*

Proof: Define

$$A = [f_1\ f_2\ \cdots\ f_{n_1}]$$

By doing a SVD of A we get

$$\begin{aligned} A &= U\Sigma V^t \\ &= [u_1\ u_2\ \cdots\ u_{mn}]\Sigma[v_1\ v_2\ \cdots v_{n_1}]^t \end{aligned}$$

where $\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_r \\ O & \cdots & \cdots & O \end{pmatrix}$ and r is the rank of A. (It may be noted that r is atmost mn.) Also it can be shown that $AA^t u_i = \sigma_i^2 u_i$, that is, σ_i^2 are the eigen values of AA^t and u_i are the corresponding eigen vectors. It can be shown that [4] $\{u_1, u_2 \cdots u_r\}$ is an orthonormal basis for the range $R(A)$ of A. If $rank(A) = r \leq n$ then the elements of C_1 are not linearly independent. By taking the canonical basis we get $A \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = f_1 \Rightarrow f_1 \in R(A)$. This can be shown for all $f_i \in C_1$. Thus $C_1 \subseteq R(A)$. Therefore W_{C_1} is an orthonormal basis for C_1. Hence

$$W_{C_1} =< u_1, u_2, \cdots, u_r >= C_1$$

Similarly define $B = [g_1\ g_2\ \cdots\ g_{n_2}]$ and $rank(B) = s$, then as above $\{v_1, v_2, \cdots, v_s\}$ is an orthonormal basis for C_2. Therefore

$$W_{C_2} =< v_1, v_2 \cdots v_s >= C_2$$

Now define $C = [A|B]$ and W_C is the orthonormal basis of $< C >$. Also $D = [W_{C_1}|W_{C_2}]$ and W_D is the orthonormal basis of $< D >$.

$$D = [u_1\ u_2\ \cdots\ u_r|v_1\ v_2\ \cdots\ v_s]$$

$f \in < D > \Rightarrow f = \Sigma_i(\alpha_i u_i + \beta_i v_i) = \Sigma_i(\alpha_i(\Sigma_j \gamma_j f_j) + \beta_i(\Sigma_j \delta_j g_j)) \Rightarrow f \in C$. Again $g \in C \Rightarrow g \in \Sigma_i(\alpha'_i f_i + \beta'_i g_i) = \Sigma_i(\alpha'_i(\Sigma_j \gamma'_j u_j) + \beta'_i(\Sigma_j \delta'_j v_j)) \Rightarrow g \in < D >$. Therefore $< C >=< D >$ and $W_{C_1} = W_{C_2}$.

When we are doing a PCA we are concerned with the projection on the range space. The contribution in terms of error from $R^{\perp}$ (complement of the

range space). By taking the orthonormal basis, that is, $\{u_1, u_2, \cdots u_m\}$ we get the entire basis for PCA . But the contribution to the reconstruction error comes from the projection to the range space alone. Let $h \notin C, h \in R^{mn} \Rightarrow h \notin D$. Let h_C be the best approximation to h in $W_C =< C >=< A, B >$. Then

$$||h - h_C|| \leq ||h - h_i|| \; \forall \; h_i \in W_C$$

Now $h_C \in W_C \subseteq < W_{C_1}, W_{C_2} >$. Let h_D be the best approximation to h in $< W_{C_1}, W_{C_2} >$. By definition of best approximation

$$||h - h_D|| \quad \leq \quad ||h - h_C|| \tag{1}$$

Also since h_C is the best approximation in $< C_1, C_2 >$,

$$||h - h_C|| \quad \leq \quad ||h - h_D|| \tag{2}$$

From 1 and 2 we get

$$||h - h_C|| = ||h - h_D||$$

This shows that the error is the same.

The two reconstructions are different although the error is the same. Either of the two can be used. They are separately unique in their respective subspaces.

In order to reduce the information, we could use a subspace of W_C and W_D. In this case the error can be mathematically characterized in the following way: $C = [A|B] = [C_1|C_2]$. Now $W_C =< u_1, u_2 \cdots u_r >$ is an orthonormal basis of C_1. This can be extended to an orthonormal basis of C. Take v_1, if it is dependent on W_C discard it else orthonormalise $\{u_1 \cdots u_r, v_1\}$. Continue till we get a $r + s$ dimensional set.

In this section, we have presented a method of establishing a class hierarchy. In the next section, we discuss the appearance based classification scheme developed using this hierarchical combination of appearance space.

3 Appearance based Classification Algorithm

In this section we describe the algorithm that we use for the purpose of classification of a given object in the hierarchy that has been made by the progressive combination of the eigen spaces corresponding to different objects. Clearly the classification here involves traversing through the hierarchy based on a decision criterion and arriving at the most probable semantic class to which that object belongs.

Given several objects, we first establish a class hierarchy on these objects. Now given a novel object, the aim is to be able to identify the position of this object in the established class hierarchy or indicate its absence. Presented below is the algorithm for Appearance Based Classification.

3.1 Algorithm

Given : A collection of views of a set of model objects and a view of an unknown object; information about hierarchical categories of the model objects
Output : Class label of the unknown object
Procedure

1. Represent each view of the model object as an *image vector* and form a set of image vectors corresponding to each object.
2. Do a PCA on each such set of image vectors and construct a hierarchy of *Appearance Classes* using the information about categories. The higher levels of the hierarchy are formed by repeatedly combining classes, as shown in Fig. 1.
3. Given a view of an unknown Object say O, we first compute the Reconstruction Error with respect to the highest class. Compute the reconstruction error with respect to all nodes at its immediate lower level.
4. If (reconstruction error(level i+1)) is slightly higher than the (reconstruction error (level i)), traverse down the hierarchy by taking the branch of which the reconstruction error is minimum among its siblings.
5. Continue this to a finer level till the Reconstruction Error values remain constant or differ very slightly.
6. Stop at the level after which the Reconstruction Error shows a sudden increase.
7. The level before this is the level to which O is to be classified.

In Theorem 2.2, we have shown that the Reconstruction Error remains unchanged between the compound appearance class formed from two elementary appearance classes or the appearance class corresponding to the compound set. For the above algorithm of Appearance Based Classification to hold, we need to further show that the reconstruction of an image vector f with respect to a space has a lower reconstruction error than the reconstruction of f with respect to any of its subspaces. That is the reconstruction error goes on increasing as we move from a space to its subspaces, that is, as we move from an appearance class to its lower appearance classes. This is shown in Theorem 3.1.

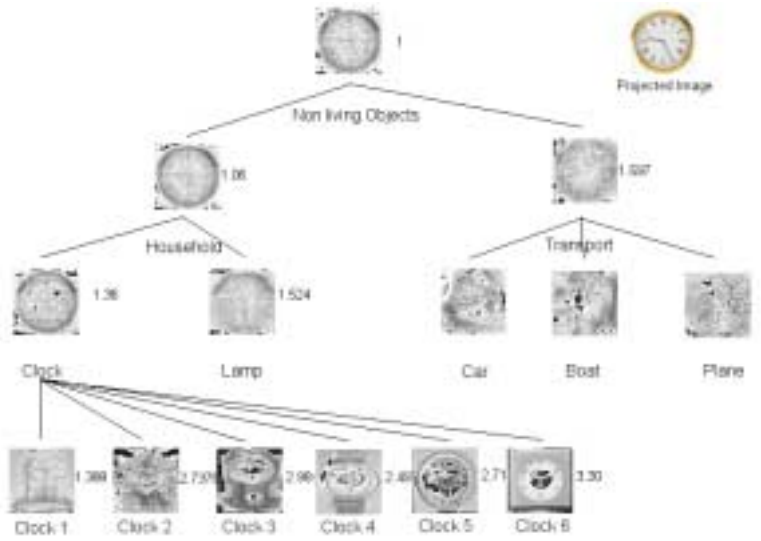

Figure 4: Results of the Classification algorithm. The reconstruction errors have been normalized with respect to the error in the root node

Theorem 3.1 *Let f be any image vector. Let C and C' be two sets and W_C and $W_{C'}$ be the corresponding appearance classes such that $C \subseteq C'$. Let f_C and $f_{C'}$ be the best reconstructions of f in W_C and $W_{C'}$ respectively. Then*

$$||f - f_{C'}|| \leq ||f - f_C||$$

Proof : Since W_C and $W_{C'}$ are orthonormal sets, then applying Theorem 2.1, we can compute the best approximation f_C and $f_{C'}$, to f respectively in the two spaces. By definition of best approximation

$$||f - f_{C'}|| \leq ||f - h_i|| \text{ for all } h_i \in W_{C'}, i = 1, 2 \cdots nm$$

Now $f_C \in W_{C'}$, therefore

$$||f - f_{C'}|| \leq ||f - f_C||$$

The above theorem shows that the reconstruction error goes on increasing as we go down the class hierarchy.

Let us consider the class C_1, C_2 such that W_C is the basis of the space generated by $C_1 \cup C_2, C_1 \cap C_2 = \phi$ and W_D is the basis of the space $W_{C_1} \cup W_{C_2}$. Let us call W_D as class B_1 and W_C as class A_1 for simplicity. If the novel Object O belongs to class B_1 or is visually similar to it, then the Reconstruction Error with respect to classes A_1 and B_1 should be the same, by Theorem 2.2. Due to noise the values may differ upto an empirically fixed threshold. The Reconstruction Error shows a sudden increase with respect to class C_1 and C_2 as is shown in Theorem 3.1. This shows that the O belongs to the class B_1. In the next section, we consider various objects and sequences and establish a class hierarchy on these. We then classify various novel objects based on their appearance.

4 Results and Discussion

In this section we present the results of the experiments carried out on the classification algorithm.

4.1 Formation of the Representation Space

In this section we show some of the generic objects (Fig. 2) that were used to build up the appearance based representation space. These representation were combined using available semantic information to build up the sample appearance based hierarchy Fig. 3. The appearance based semantic grouping is a manual decision.

4.2 Classification of Objects

In figure 4 the performance of the classification algorithm has been shown with the test images of a novel clock. As can be seen from the figure, the reconstruction error increases as we traverse down the hierarchy. Our decision criterion is to select the node for which the reconstruction error is minimum as compared to the reconstruction errors for other nodes at the same level in the hierarchy (siblings). Moreover we terminate this process at the level i when the error at the *(i+1)th* is greater than 1.5 times the error at the *ith*

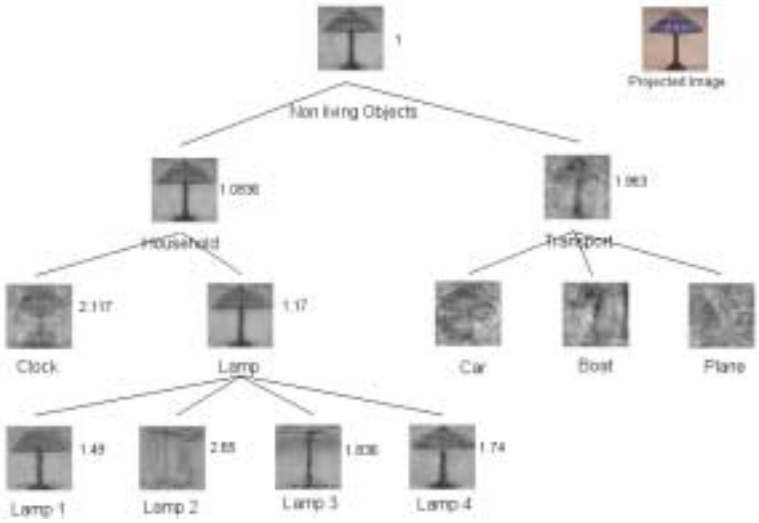

Figure 5: Another result of the classification algorithm

level. As we can see from the figure 4, the classification was terminated at class level of clocks since the clock that had been queried was a novel clock which did not actually belong to any of the clock models that had been used in populating the database. Therefore the classification algorithm has been correctly able to determine the class of the given clock. Similar performance of classification is also obtained with the lamp example where the algorithm has been able to correctly determine the model to which the lamp belongs (as shown in figure 5.

In figure 6 we present the overall performance of the classification algorithm. The tests were carried out for over 140 test images and correct classification of the object could be achieved in about 95% of the cases. The misclassifications in the case of some query images could be attributed to the fact that these query images differed entirely from the objects in the database in terms of their shapes and appearances. The car and boat, in some cases, are misclassified due to their visual similarity which could be confusing to humans too. This indicates the strength of the appearance based classification. Colour and patterns on the object do not effect the visual similarity of the objects and hence in most cases , in spite of difference in colour, we have achieved correct classification. The classification we have achieved in the case of the 140 images in the database indicate the applicability of the scheme for classification. This problem can however be significantly alleviated by making the database more comprehensive.

5 Conclusion and Scope for Future Work

In this paper we have proposed an algorithm for collection and representation of generic objects. This involves the formation of a compact eigen space based representation of generic object. We have then proposed to use this representation for the purpose of building up an appearance class based hierarchy wherein the objects that belong to the same class are grouped together and their respective eigen spaces are progressively combined to form the appearance class corresponding to that group of objects. We, have further shown that the appearance classes formed by such a progressive combination are the optimal representations of the classes to which the training examples belong. Experimental results clearly show that the system based on the above algorithm correctly determines the class to which an object belongs thereby demonstrating the correctness of the proposed algorithm.

This work therefore provides the basis for the development of an object database that can be used for supporting content based querying. This approach greatly enhances the functionality of such a retrieval system because of its ability to resolve unknown queries at an approximate level that can then be returned to the user for further refining of the query. Another interesting application of such a hierarchical representation of objects is in solving the problem of object recognition reliably. Often recognition algorithms fail because the invariants they use do not

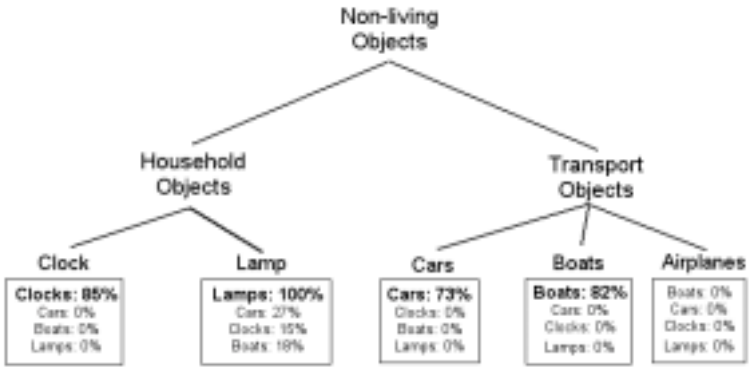

Figure 6: Classification Statistics. The percentages represent the ratio of the query images of a particular class that were classified at that particular level in the hierarchy

in general hold for a large variety of objects and are rather valid for particular classes of objects. This kind of a representation of objects can therefore be used as the first step for arriving at the approximate class of an object after which class specific invariants can be applied for accurate recognition.

References

[1] S. Baker and S.K. Nayar, "Pattern Rejection", Proceedings of the 1996 IEEE Conference on Computer Vision and Pattern Recognition, pp. 544-549, June, 1996.

[2] T. F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham, "Active Shape Models - their training and application", Computer Vision Graphics and Image Understanding 61 (1), pp. 39-59, 1995.

[3] S. Duvdevani-Bar and S. Edelman, "Visual recognition and categorization on the basis of similarities to multiple class prototypes", International. Journal of Computer Vision 33(3),pp. 201-228, 1999.

[4] K. Hoffman and R. Kunze, "Linear Algebra", Prentice-Hall, 1971.

[5] H. Murase and S. K. Nayar, "Visual Learning and Recognition of 3D Objects from Appearance," International Journal of Computer Vision, Vol. 14(1),pp. 5-24, 1995.

[6] B. Schiele and James L. Crowley, "The concept of Visual Classes for Object Classification", Scandinavian Conference on Image Analysis, Lappeenranta, Finland, June 1997.

[7] L. Stark and K. Bowyer, "Generic Object Recognition using Form and Function", World Scientific Press, 1996.

[8] M. Turk and A. Pentland, "Eigenfaces for recognition", Journal of Cognitive Neuroscience 3(1), pp. 71-86, 1991.