View-Invariant Recognition Using Corresponding Object Fragments

Evgeniy Bart, Evgeny Byvatov, and Shimon Ullman

Department of Computer Science And Applied Mathematics
Weizmann Institute of Science, Rehovot, Israel 76100
{evgeniy.bart, shimon.ullman}@weizmann.ac.il

Abstract. We develop a novel approach to view-invariant recognition and apply it to the task of recognizing face images under widely separated viewing directions. Our main contribution is a novel object representation scheme using 'extended fragments' that enables us to achieve a high level of recognition performance and generalization across a wide range of viewing conditions. Extended fragments are equivalence classes of image fragments that represent informative object parts under different viewing conditions. They are extracted automatically from short video sequences during learning. Using this representation, the scheme is unique in its ability to generalize from a single view of a novel object and compensate for a significant change in viewing direction without using 3D information. As a result, novel objects can be recognized from viewing directions from which they were not seen in the past. Experiments demonstrate that the scheme achieves significantly better generalization and recognition performance than previously used methods.

1 Introduction

View-invariance refers to the ability of a recognition system to identify an object, such as a face, from any viewing direction, including directions from which the object was not seen in the past. View-invariant recognition is difficult because images of the same object viewed from different directions can be highly dissimilar. The challenge of view-invariant recognition is to correctly identify a novel object based on a limited number of views, from different viewing directions. For example, after seeing a single frontal image of a novel face, the same face has to be recognized when seen in profile.

In the current study we develop a scheme for view-invariant recognition based on the automatic extraction and use of corresponding views of informative object parts. The approach has two main components. First, objects within a class, such as face images, are represented in terms of common 'building blocks', or parts. The parts we use are sub-images, or object fragments, selected automatically from a training set during a learning phase. Second, images of the same part under different viewing directions are grouped together to form a generalized fragment that extends across changes in the viewing direction. (We therefore refer to a set of equivalent fragments as an 'extended' fragment.) The general

T. Pajdla and J. Matas (Eds.): ECCV 2004, LNCS 3022, pp. 152–165, 2004.

idea is that the equivalence between different object views will be induced by the learned equivalence of the extended fragments. To achieve view invariance, the view of a novel object within a class will be represented in terms of the constituent parts. The appearances of the parts themselves under different conditions are extracted during learning, and this will be used for recognizing the novel object under new viewing conditions. We describe below how such extended fragments are extracted from training images, and how they are used for identifying novel objects seen in one viewing direction, from a different and widely separated direction.

The remainder of the paper is organized as follows. In section 2 we review past approaches to view-invariant recognition. In section 3 our extended fragments representation is introduced, and in section 4 we describe how this representation is incorporated in a recognition scheme. In section 5 we present results obtained by our algorithm and compare them to a popular PCA-based approach. We make additional comparisons and discuss future extensions in section 6.

2 A Review of Past Approaches

In this section, we give a brief review of some general approaches to view-invariant recognition.

One possible approach to achieve view invariance is to use features that are by themselves invariant to pose transformations. The basic idea is to identify image-based measures that remain constant as a function of viewing direction, and use them as a signature that identifies an object. One well-known measure is the four-point cross-ratio, but other, more complicated algebraic invariants, have been proposed [1]. Several types of features invariant under arbitrary affine transformations were derived and used for object recognition [2,3], and features that are nearly invariant were derived for more general transformations [4]. One shortcoming of this approach is that it is difficult to find a sufficient number of invariant features for reliable recognition, especially when objects that are similar in overall shape (such as faces) have to be discriminated. Second, many useful features are not by themselves invariant, and consequently their use is excluded in the invariant features framework, in contrast with the extended features approach described below . (See also comparisons in section 3.)

Another general approach is to store multiple views of each object to be recognized, and possibly apply some form of view-interpolation for intermediate views (e.g. [5]). This approach requires multiple views of each novel object, and the interpolation usually requires correspondence between the novel object and a stored view. Such correspondence turned out in practice to be a difficult problem.

Having a full 3D model of an object alleviates the need to store multiple views, since novel views may be generated from such a model. However, obtaining precise 3D models in practice is difficult, and usually requires special measuring equipment (e.g. [6]). Due to this requirement, recognition using 3D data is frequently considered separately from image-based methods. For examples of 3D approaches, see [7,8].

Several methods (elastic graph matching [9], Active Appearance Models [10]) use flexible matching to deal with the deformation caused by changes in pose.

Since small pose changes tend to leave all features visible and only change the distances between them, this approach is able to compensate for small ($10 - 15°$) head rotations. A feature common to such approaches is that they easily compensate for small pose changes, but the performance drops significantly when larger pose changes (e.g. above $45°$) are present.

A popular approach to object recognition in general is based on principal components analysis (PCA). When applied to face recognition, this approach is known as eigen-faces [11]. Several researchers have used PCA to achieve pose invariance. Murase and Nayar [12] acquired images of several objects every four degrees. From these images, they constructed an eigenspace representation for a given object, and used it for recognizing the object in different poses. A limitation of this approach is the need to acquire and store a large number of views for each object. Pentland et al. [13] developed a similar scheme, applied to face images, and using only five views between frontal and profile (inclusive). Performance was good when the view to be recognized was at the same orientation as the previously seen pictures, but dropped quickly when interpolation or extrapolation between views was required.

3 The Extended Fragments Approach

Our approach is an extension of object recognition methods in which objects are represented using a set of informative sub-images, called fragments or patches [14,15,16]. These methods are general and can be applied to a wide variety of natural object classes (for example, faces, cars, and animals). In this section we describe briefly the relevant aspects of the fragment-based approaches, discuss their limitations for view-invariance, and outline our extension based on extended fragments. We illustrate the approach using the task of face recognition, but the method is general and can be applied to different object classes.

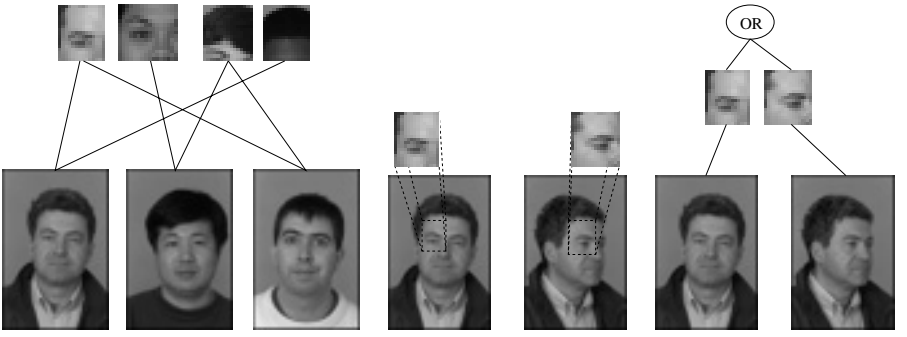
In fragment-based recognition, informative object fragments are extracted during a learning stage. The extraction is based on the measure of mutual information between the fragments and the class they represent. A large set of candidate fragments is evaluated, and a subset of informative fragments is then selected for the recognition process. Informative fragments for face images typically include different types of eyes, mouths, hairlines, etc.

During recognition, this set of fragments is searched for in the target images using the absolute value of normalized cross-correlation, given by

$$NCC(p, f) = \frac{E(p - \overline{p})(f - \overline{f})}{\sigma_p \sigma_f}. \quad (1)$$

Here f is the fragment and p is an image patch of the same size as the fragment. Image patches at all locations are evaluated and the one with the highest correlation is selected. When the correlation exceeds a pre-determined threshold, the fragment is considered present, or *active*, in the image. A schematic illustration of this scheme is presented in Figure 1(a).

Informative fragments have a number of desirable properties [14]. They provide a compact representation of objects or object classes and can be used for



(a) Fragments-based recognition (b) Lack of viewpoint-invariance (c) Extended fragments

Fig. 1. Extended object fragments. (a) Schematic illustration of previous fragments-based approaches. Bottom: faces represented in the system. Top: informative fragments used for the representation. Lines connect each face to fragments that are present in the face, as computed by normalized cross-correlation. Novel faces can be detected reliably using a limited set of fragments. (b) Informative fragments are not viewpoint-invariant, for instance, a frontal eye fragment (left) is different from the corresponding side fragment (right). If the detection threshold is set low enough so that the fragment will be active in both images, many spurious detections will occur, and the overall recognition performance will deteriorate. (c) View invariance is obtained by introducing equivalence sets of fragments. Fragments depicting the same face part viewed from different angles are grouped to form an extended fragment. The eye is detected in an image if either the frontal or the profile eye templates are found. This is indicated by the OR attached to the pair of fragments.

efficient and accurate recognition. However, fragments used in previous schemes are not view-invariant. The reason is that objects and object parts look very different under different orientations. As a result, fragments that were active e.g. in a frontal view of a certain face will not be active in side views of the same face. Therefore, the representation by active fragments is not view-invariant. This problem is illustrated in Figure 1(b).

To overcome this problem, we use the fact that the only source of difference between the left and right images in Figure 1(b) is different viewpoint. The face itself consists of the same sub-parts in both images, and we therefore wish to represent the objects in terms of sub-parts and not in terms of view-specific sub-images. The representation using sub-parts will then be view-invariant and will allow invariant recognition.

To represent sub-parts in an invariant manner, one approach has been to use affine-invariant patches [3,2]. This approach works well in some applications (such as wide-baseline matching) that use nearly planar surfaces. However, for non-planar objects, including faces, affine transformations provide a poor appro-

ximation. In our experiments, methods based on affine invariant matching failed entirely at 45° rotation.

In our scheme, invariance of sub-parts was achieved using multiple templates, by grouping together the images of the same object part under different viewing conditions. For example, to represent the 'eye' part in Figure 1(b), the two sub-images of the eye region shown in the figure are grouped together to form an 'extended eye fragment'. Using this extended fragment, the eye is detected in an image if either the frontal or the side eye template are detected. Typically, the frontal template would be found in frontal images and the profile template would be found in profile images. Consequently, at the level of extended fragments, the representation becomes invariant, as illustrated schematically in Figure 1(c).

Note that the scheme uses only multiple-template representation for *object parts*, not for entire objects as was used by previous multiple template algorithms [5,12,13]. This has a number of significant advantages over previous schemes. First, multiple examples of each object are needed only in the training phase, when extended fragments are created. Since extended fragments are both view-invariant and capable of representing novel objects of the same class, view-invariant representation of novel objects is obtained from a single image. This is a significant advantage over previous multiple-views schemes, where many views for each novel objects were required. Second, the extended fragments representation is more efficient in terms of memory than previous multiple-template schemes, because the templates required for fragment representation are much smaller that the entire object images. The reduction in space requirements also reduces matching time, as there is no need to perform matching of a large collection of full-size images. Finally, object parts generalize better to novel viewing conditions than entire objects (see section 6). Therefore, fewer templates per extended fragment will be required to cover a given range of viewing directions compared with matching images of entire objects.

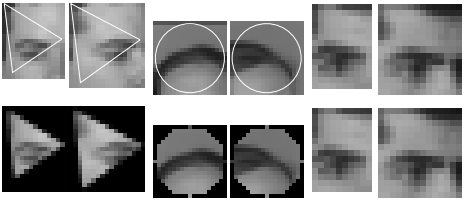
We have implemented the scheme outlined above and applied it to recognize face images from two widely separated viewing directions: frontal and 60° profile, called below a 'side view'. Note that this range is wide enough to undermine schemes that were not specifically designed for view-invaraince, such as [9,10]. As discussed in section 6, generalization of our algorithm to handle any viewing direction is straightforward. The following sections describe in more detail the different stages in the algorithm.

3.1 The Extraction of Extended Fragments

In our training, extended fragments were extracted from images of 100 subjects from the FERET database [17]. The images were low-pass filtered and down-sampled to size 60×40 pixels.

To form extended fragments, the multiple template representation of object parts must be obtained. To deal with two separate views, a set of sub-image pairs must be provided, where in each pair one sub-image will be a view of some face part in frontal orientation and the other sub-image will be a view of the same face part in the side view. For this, correspondence must be established between

(a) A sample training sequence



(b) Example fragments

Fig. 2. Illustration of extended fragments selection. (a) A sample training sequence. (b) Some extended fragments that were selected automatically from this and similar training sequences. Each pair constitutes an extended fragment; the left part is the fragment in frontal images, the right part is the same fragment in side images. Top part: sub-images with fragment shapes delineated. Bottom part: the same fragments are shown outside the corresponding sub-images.

face areas in frontal and side images. This is a standard task in computer vision that is similar to optical flow computation.

We used the KLT (Kanade-Lucas-Tomasi) algorithm [18] to automatically establish correspondences between frontal and side views. The KLT algorithm selects points in the initial image that can be tracked reliably, and uses a simple gradient search to follow the selected points in subsequent images. The tracking improves when the differences between successive images are small. Therefore, using short video segments of rotating faces produces more reliable results. We have used in the training stage the images from the FERET database [17]. A training sequence contained three intermediate views in addition to the frontal and size views; an example of such a sequence is shown in Figure 2(a).

After tracking was completed, the intermediate views were discarded. The correspondences obtained by KLT can then be used to associate together the views of the same face part under different poses. This is obtained by selecting a sub-image in one view (e.g. frontal), and using the tracked points to identify the corresponding sub-image from the other (side) view. The sub-images are polygons defined by a subset of matching points. In particular, we have used triangular sub-images defined by corresponding triplets of points. Matching pairs of triangles (one from frontal, another from side view) were grouped together and formed the pool of candidate extended fragments. We have also tried to interpolate between the points tracked by KLT to obtain dense correspondences, and use matching regions of arbitrary shape (Figure 2(b)). However, the difference in performance was only marginal. Therefore, triangular fragments were used throughout most of our tests.

3.2 Selection Using Mutual Information and Max-Min Algorithm

During learning, extended fragments were extracted from the 100 training sequences. The number of all possible fragments was about 100000 per sequence;

as a result, learning becomes time-consuming. However, most of these candidate fragments are not informative, and it is possible to reduce the size of the pool based on the fragments size. It was shown in [15] that most informative fragments have intermediate size. In our experiments (see section 6), fragments that were smaller than 6% of the object area or larger than 20% were uninformative. By excluding from consideration candidate fragments outside these size constraints, the number of candidate extended fragments was reduced to about 1000 per training sequence. Since calculating the fragment's size is much simpler than evaluating its mutual information, significant savings in computation were obtained.

Extracting extended fragments from all 100 sequences results in a pool of candidate fragments of size around 100000. This set still contains many redundant or uninformative extended fragments. Therefore, the next stage in the feature extraction process is to select a smaller subset of fragments that are most useful for the task of recognition.

This selection was obtained based on maximizing the information supplied by the extended fragments for view-invariant recognition. The use of mutual information for feature selection is motivated by both theoretical and experimental results. Successful classification reduces the initial uncertainty (entropy) about the class. The classification error is bounded by the residual entropy (Fano's inequality [19]), and this entropy is minimized when $I(C;F)$, the mutual information between the class and the set F of fragments, is maximal. In practice, selecting features based on maximizing mutual information leads to improved classification compared with less informative features [14]. We explain below the procedure for selecting the most informative extended fragments.

This first step of the selection procedure derives for each extended fragment a measure of mutual information. Mutual information between the class C and fragment F is given by

$$I(C;F) = \sum_{c,f} p(C=c, F=f) \log \frac{p(C=c, F=f)}{p(C=c)p(F=f)}. \tag{2}$$

By measuring the frequencies of detecting F inside different classes c, we can evaluate the mutual information of a fragment from the training data.

The next step is to select a bank of n fragments $B = \{F_1, \ldots, F_n\}$ with the highest mutual information about the class C; $B = \arg\max I(C;B)$. Evaluating the mutual information with respect to the joint distribution of many variables is impractical, therefore some approximation must be used. A natural approach is to use greedy iterative optimization. The selection process is initialized by selecting the extended fragment F_1 with the highest mutual information. Fragments are then added one by one, until the gain in mutual information is small, or until a limit on the bank size n is reached. To expand a size-k fragment bank $B = \{F_1, \ldots, F_k\}$ to size $k+1$, a new fragment F_{k+1} must be selected that will add the maximal amount of new information to the bank. The conditional mutual information between F_{k+1} and the class given the current fragment bank must therefore be maximized: $F_{k+1} = \arg\max I(C; F_{k+1}|B)$. Estimating $I(C; F_{k+1}|B)$ still depends on multiple fragments. The term $I(C; F_{k+1}|B)$ can be approximated by $\min_{F_i \in B} I(C; F_{k+1}|F_i)$. This term contains just two fragments and can

be computed efficiently from the training data. The approximation essentially takes into account correlations between pairs of fragments, but not higher order interactions. It makes sure that the new fragment F_{k+1} is informative, and that the information it contributes is not contained in any of the previously selected fragments. The overall algorithm for selection can be summarized as:

$$F_1 = \arg\max_F I(C; F); \tag{3}$$

$$F_{k+1} = \arg\max_F \min_i I(C; F|F_i). \tag{4}$$

The second stage determines the contribution of a fragment F by finding the most similar fragment already selected (this is the min stage) and then selects the new fragment with the largest contribution (the max stage). The full computation is therefore called the max-min selection.

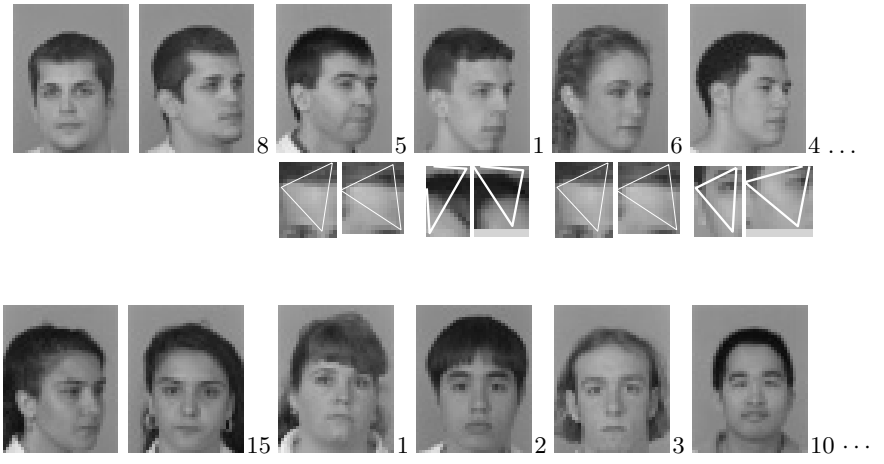


Fig. 3. An example of recognition by extended fragments. (a) Top row: a novel frontal face (left image), together with the same face, but at 60° orientation, among distractor faces. Only a few examples are shown, the actual testing set always contained 99 distractors. The side view images are arranged according to their similarity to the target image computed by the extended fragments algorithm. The image selected as the most similar is the correct answer. Below each distractor, one of the extended fragments that helped in the identification task is shown. Each of these fragments was detected either only in the frontal face, or only in the distractor side view above the fragment, providing evidence that the two faces are different. The numbers next to each face show its rank as given by the view-based PCA scheme. (b) Same as (a), without the fragments shown. The test faces were frontal in this case and the target face was at 60°.

During recognition, fragment detection is performed by computing the absolute value of the normalized cross-correlation at every image location, and

selecting the location with the highest correlation. The maximal correlation, which is a continuous value in the range $[0, 1]$, can be used in the recognition process. We used, however, a simplified scheme in which the feature value was binarized. A fragment was considered to be present, and have the value of 1 in a given image, if its maximal correlation in the image was above a pre-determined threshold. If the maximal correlation was below the threshold, the fragment was assigned value of 0. (Fragments whose activation is above the threshold are also called 'active' below, and those with activation below the threshold are called 'inactive'.) An optimal threshold was selected automatically for each fragment in such a way as to maximize the fragment's mutual information with the class: $\theta = \arg\max I(C; F_\theta)$. Here F_θ is the fragment F detected with threshold θ. Since extended fragments consist of several individual fragments (two in our case), each has a separate threshold. These thresholds can be determined by a straightforward search procedure.

Examples of extended fragments selected automatically by the algorithm are shown in Figure 2(b).

4 Recognition Using Extended Fragments

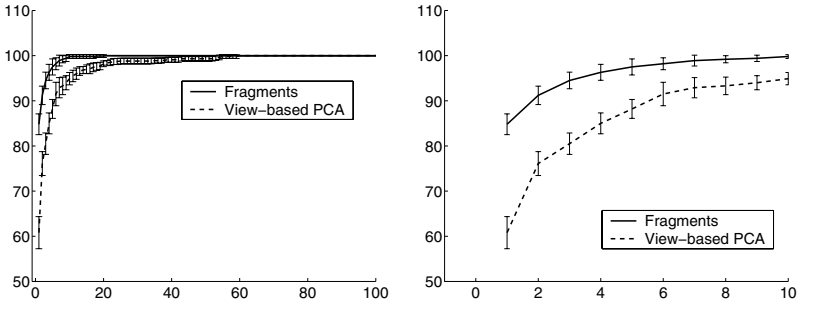
In performing recognition, the system is given a single image of a novel face, for example, in frontal view. It is also presented with a gallery of side views of different faces. The task is to identify the side view image from the gallery that corresponds to the frontal view.

Given the extended fragments representation, the recognition procedure is straightforward. The novel image is represented by the activation pattern of the fragment bank. This is a binary vector that specifies which of the extended fragments were active in the image. Similarly, activation patterns of the gallery images are known. SVM classifier was used to identify the side activation pattern that corresponds to the given frontal activation pattern. An example of the scheme applied to target and test images is shown in Figure 3.

5 Results

In this section we summarize the results obtained by the method presented above and compare them with other methods. The results were obtained as follows. The database images were divided into a training and a testing set (several random partitions were tried in every experiment). In the training phase, images of 100 individuals were used. For each individual the data set contained 5 images in the orientations shown in Figure 2(a). This set of images was used to select 1000 extended fragments and their optimal thresholds as described in section 3.2 and train the SVM classifier.

In the testing phase, the algorithm was given a novel frontal view, called the target view. (All individuals in the testing and training phases were different.) The task was then to identify the side view of the individual shown in the novel frontal view. The algorithm was presented with a set (called the 'test set') of 100 side views of different people, one of which was of the same individual shown in



(a) Recognition performance (b) Same as (a), magnified

Fig. 4. Recognition results and some comparisons. The graph value at $X = k$ shows the percentage of trials for which the correct classification was among the top k choices. Bars show standard deviation. (a) Comparison of extended fragments with PCA. (b) Initial portion of the plot in (a), magnified.

the frontal view. The algorithm ranked these pictures by their similarity to the frontal view using the extended features as described above. When the top ranked picture corresponded to the target view, the algorithm correctly recognized the individual.

We present our results using CMC (cumulative match characteristics) curves. A CMC curve value at point $X = k$ shows the percentage of trials for which the correct match was among the top k matches produced by the algorithm. Typically, the interesting region of the curve comprises the several initial points; in particular, the point $x = 1$ on the curve corresponds to the frequency at which the correct view was ranked first among the 100 views, i.e. was correctly recognized.

The results of our scheme are shown in Figure 4. As can be seen, side views of a novel person were identified correctly in about 85% of the cases following the presentation of a single frontal image. In order to compare this performance to previous schemes, we implemented the view-invariant PCA scheme of Pentland et al. [13], which is one of the most successful and widely used face recognition approaches. Our implementation was identical to the scheme as described in [13]. PCA performance was calculated under exactly the same conditions as used for our algorithm, i.e. we trained PCA on the same training images and tested recognition on the same images. Figure 4(a) shows the results of the comparison. As can be seen from the figure, this method identifies the person correctly in 60% of the cases. The plots in Figure 4 show the marked advantage of the present algorithm over PCA (the differences are significant at $p < 0.01$, χ^2 test).

A recognized weakness of the PCA method is that it requires precise alignment of the images. In contrast, our algorithm can tolerate significant errors in alignment. We have tested the sensitivity of both algorithms to alignment precision. To test the sensitivity of the extended fragments scheme, we fixed one

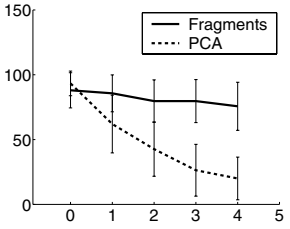


Fig. 5. The effect of misalignment on recognition performance. Percent correct recognition as a function of misalignment magnitude in pixels for extended fragments and view-invariant PCA.

(frontal) part of each fragment, and shifted the other (side) part in a random direction by progressively larger amounts from its correct position. This created a controlled error in the location of corresponding fragments. We tested the recognition performance as a function of the correspondence error. The results were compared with a similar test applied to the PCA method, where the images were not precisely aligned, but had a systematic misalignment error. Figure 5 shows the performance of the schemes as a function of the amount of misalignment. (These tests were performed on a new database and with a smaller test set, therefore the results are not on the same scale as in previous figures.) The task was to recognize one out of five faces. Note that for four-pixel shifts, corresponding on average to 12% of the face size, PCA performance reduces to chance level. In many schemes, image misalignment during learning is a significant potential source of errors. As seen in the figure, extended fragments are significantly more robust than PCA to misalignments in the learning stage.

6 Discussion

We described in this work a general approach to view-invariant object recognition. The approach is based on a novel type of features, which are equivalence sets of corresponding sub-images, called extended fragments. The features are class-based and are applicable to many natural object classes. In particular, we have applied the approach to cars and animals with similar results.

Despite the large number of extended fragments used, time requirements of the recognition stage are quite reasonable. In our tests, 1500 recognition attempts using 1000 fragments took 2–3 seconds (without optimizing the code for efficiency). Time requirements of the learning stage are more significant, but learning is performed off-line.

One potential limitation of the extended fragments representation is that due to the local nature of the fragments, they might be detected in face images under inconsistent orientations (e.g. frontal fragment would be detected in profile view), which might lead to a decrease in performance. However, in experiments where fragments were restricted to be detected only in the relevant orientations, no performance improvement was observed. The implication is that reliable

recognition can be based on the activation of the features themselves, without explicitly testing for view consistency between different features.

In the feature selection process, the size of preferred features was not fixed, but free to change within a general range (set in the simulations between 6–20% of object size). In other schemes, features are often required to be of a fixed size, with some schemes using small local features, and others using global object features. It therefore became of interest to test the size of the most informative features, and to compare the results with alternative approaches.

Previous studies [15] have shown that useful fragments are typically of intermediate size. To investigate this further, we tested the effect of feature size on view-invariant recognition, using a new database of size 50; 40 images for training and 10 for testing. This size of the database allowed selecting extended fragments of various sizes and comparing their performance. We found that when the extended fragments were of size between 6% and 20% of the object area, the test faces were correctly recognized in $89 \pm 10\%$ of the cases. When the selected fragments were constrained to be smaller (below 3% of the object size), the performance dropped to $67 \pm 12\%$, and when the fragments were constrained to be large (above 20%), the performance dropped to $75 \pm 10\%$. Both results were highly significant (t test, $p < 0.01$). This can be contrasted with approaches such as [20,21], that use small local features, or [11,13], where the features are global. The optimal features extracted were also found to vary in size, to represent object features of different dimensions. This can be contrasted with approaches such as [22], where the features are constrained to have a fixed size. This flexibility in feature size is important for maximizing the fragments mutual information and classification performance.

In our testing, features were extracted for two orientations – frontal and 60° side view. A complete recognition model should be able, however, to handle a full range of orientations (from left to right profile and at different elevations). As mentioned above, the scheme can be extended to handle a full range of orientations by including views from a set of representative viewing directions in each extended fragment. Results of a preliminary experiment suggest that 15 views are sufficient to cover the entire range of views, or nine views if the bilateral symmetry of the face is used. These requirements can be compared with typical view interpolation schemes such as [5]. This scheme requires the use of 15 views to achieve recognition within a restricted range of $-30°$ to $30°$ horizontally and $-20°$ to $20°$ vertically. More importantly, it requires all 15 views for each novel face. In contrast, the extended fragments scheme requires 15 views for training only; in testing, recognition of novel faces is performed from a single view.

The proposed approach can be extended to handle sources of variability such as illumination and facial expression. This can be performed within the general framework by adding the necessary templates to each extended fragment. There are indications [23] that compensating for illumination changes will be possible using a reasonable number of templates. Facial expressions often involve a limited area of the face, and therefore affect only a small number of fragments. The full size of equivalence sets of extended fragments required to perform unconstrained recognition is a subject for future work.

Acknowledgments. This research was supported in part by the Moross Laboratory at the Weizmann Institute of Science. Portions of the research in this paper use the FERET database of facial images collected under the FERET program [17].

References

1. Mundy, J., Zisserman, A.: Geometric Invariance in Computer Vision. The MIT press (1992)
2. Tuytelaars, T., Gool, L.V.: Wide baseline stereo matching based on local, affinely invariant regions. In: British Machine Vision Conference. (2000) 412–425
3. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: Proceedings of the European Conference on Computer Vision. (2002) 128–142
4. Wallraven, C., Bülthoff, H.H.: Automatic acquisition of exemplar-based representations for recognition from image sequences. In: CVPR 2001 - Workshop on Models vs. Exemplars. (2001)
5. Beymer, D.J.: Face recognition under varying pose. Technical Report AIM-1461, MIT Artificial Intelligence Lab (1993)
6. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In Rockwood, A., ed.: Siggraph 1999, Computer Graphics Proceedings, Los Angeles, Addison Wesley Longman (1999) 187–194
7. Nagashima, Y., Agawa, H., Kishino, F.: 3D face model reproduction method using multi view images. Proceedings of the SPIE, Visual Communications and Image Processing '91 **1606** (1991) 566–573
8. Lowe, D.G.: Three-dimensional object recognition from single two-dimensional images. Artificial Intelligence **31** (1987) 355–395
9. Wiskott, L., Fellous, J.M., Krüger, N., von der Malsburg, C.: Face recognition by elastic bunch graph matching. In Jain, L.C., Halici, U., Hayashi, I., Lee, S.B., eds.: Intelligent Biometric Techniques in Fingerprint and Face Recognition. CRC Press (1999) 355–396
10. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. In: Proceedings of the European Conference on Computer Vision. (1998) 484–498
11. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience **3** (1991) 71–86
12. Murase, H., Nayar, S.: Visual learning and recognition of 3-d objects from appearance. International Journal of Computer Vision **14** (1995) 5–24
13. Pentland, A., Moghaddam, B., Starner, T.: View-based and modular eigenspaces for face recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA (1994)
14. Sali, E., Ullman, S.: Combining class-specific fragments for object recognition. In: British Machine Vision Conference. (1999) 203–213
15. Ullman, S., Vidal-Naquet, M., Sali, E.: Visual features of intermediate complexity and their use in classification. Nature Neuroscience **5** (2002) 682–687
16. Agarwal, S., Roth, D.: Learning a sparse representation for object detection. In: Proceedings of the European Conference on Computer Vision. (2002) 113–127
17. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.: The FERET database and evaluation procedure for face recognition algorithms. Image and Vision Computing **16** (1998) 295–306

18. Tomasi, C., Kanade, T.: Detection and tracking of point features. Technical Report CMU-CS-91-132, Carnegie Mellon University (1991)
19. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley (1991)
20. Amit, Y., Geman, D.: Shape quantization and recognition with randomized trees. Neural Computation **9** (1997) 1545–1588
21. Mel, B.W.: SEEMORE: Combining color, shape and texture histogramming in a neurally-inspired approach to visual object recognition. Neural Computation **9** (1997) 777–804
22. Weber, M., Welling, M., Perona, P.: Towards automatic discovery of object categories. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. (2002) 101–109
23. Basri, R., Jacobs, D.W.: Lambertian reflectance and linear subspaces. IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003) 218–233