Mosaic Image Method: a Local and Global Method

Li Zhao Yee-Hong Yang
Computer Vision & Graphics Lab
Department of Computer Science
University of Saskatchewan
Saskatoon
CANADA, S7N 5A9

Abstract

In this paper, a new method to compute eigenimages in principal component analysis (PCA) based vision systems is presented. It is called Mosaic Image Method. In this method, the object is represented as a collection of features and their relative positions (topology). This is a local and global method. Although this method is created to account for the occlusion problem, it is found that this is a better representation in general than the traditional optimum representation. A simple algorithm based on the new representation is proposed for recognition. Thorough experiments are conducted. More than 110,000 test images with different degree of occlusion are used to test the proposed method. The new method can accommodate up to 53% occluded parts with a more than 95% correct recognition rate. To the authors' knowledge, this is the best result in the presence of occlusion in PCA-based vision systems.

Keywords: image representation, occlusion, object recognition, principal component analysis.

1 Introduction

Occlusion is a common phenomenon in real life. When an object is partially occluded by another object, it is difficult to recognize. (Obviously when the most part of the object is occluded, it is impossible to recognize it because little information on it is available.) Other optical phenomena such as shadow and specular highlight can also be regarded as occlusion since they usually change the objects' appearances drastically, and, also due to their local properties.

Recently, a method based on the traditional principal component analysis has seen its revival in computer vision. It has been used to recognize very complex objects such as the human face [13, 1, 10, 2]. But nearly all research on PCA-based vision systems treats objects as a whole. Therefore occlusion poses a major problem for these studies.

In this paper, a new scheme to compute eigenimages and a method using this new representation for recognition and reconstruction are presented. The method is called Mosaic Image Method. In this method, an image is processed as a collection of small mosaic images. It is the authors' argument that an object is feature-based and that features together with their relative positions form a representation of the *whole* object. This representation is both local and global. Thorough experiments are conducted to verify this new representation and the accompanying recognition method which is based on this new representation. To the authors' surprise, it turns out that although this new representation is created to account for the occlusion problem, the new representation is a better representation in general than the optimum representation. The images reconstructed by this new representation are much better than those by traditional PCA method. For the recognition, the proposed method can accommodate up to 53% occluded areas with a more than 95% correct recognition rate. The recognition algorithm is simple, much information is not used. Hence it is expected that more sophisticated algorithms such as neural networks can boost the recognition rate easily. To the authors' knowledge, the results are already the best in the presence of occlusion in PCA-based vision systems.

This paper is organized as follows. In the following section, an analysis of why occlusion poses a problem to current PCA-based vision systems is presented. In the same section a review of related work is given. The following section presents the Mosaic Image Method. The advantage of the new representation is verified by thorough experiments, and the results follow. Then, an Euclidean-norm-based simple recognition algorithm and a simple reconstruction algorithm are presented. The experimental results are presented in the following section. More than 110,000 test images are used to test the proposed method. The paper concludes with a summary and a discussion of further research.

2 Why Occlusion Is a Problem in Traditional PCA-Based Vision Systems

In traditional PCA-based vision systems, eigenimages are generated by processing an image as a whole and by solving the corresponding eigensystem. Consequently, in the covariance matrix, there is an element $COV(x)_{i,j}$ for any two pixels in the image, i.e., to compute the principal components it has to account for the correlation of every pair of pixels in the image. This not only makes the computation very expensive, it also makes this method sensitive to occlusion because it depends on the whole image. Furthermore, it is hard to justify that every pixel in the image correlates to each other or that this correlation is relevant.

In the recognition stage, when a new input image (denote as $\mathbf{I}$) is presented to the vision system (assume the segmentation is done), the mean image $\mathbf{m}$ is subtracted from the input image $\mathbf{I}$. The resulting image (denoted as $\tilde{\mathbf{I}}$) is projected to the principal components. The projections and the residual values are used in the recognition stage. Now consider what will happen if the input image has some occluded parts.

In the presence of occluded parts, the image becomes $\mathbf{I} + \Delta$ (Δ is the difference image between the images with and without occluded parts). If most of the object is occluded, the object cannot be recognized because no information on the object is available. Therefore it is reasonable to assume that at least some elements of the vector Δ are zero. Assume that the principal components are Φ_i, $(i = 0, 1, 2, ..., t-1)$ (t is the number of the principal components used in the vision system). These principal components are mutually orthogonal unit vectors. The projections of the occluded image to the principal components are as follows,

$$(\mathbf{I} + \Delta)\Phi_i = \mathbf{I}\Phi_i + \Delta\Phi_i. \tag{1}$$

Here $i = 0, 1, 2, ..., t-1$. The correct projections of the non-occluded object are $\mathbf{I}\Phi_i$. $\Delta\Phi_i$ are errors introduced by the occluded parts. But the problem here is that the projections are *scalar* values. Therefore it is not possible to separate the correct projections from the errors. If these projections are used to evaluate which training sample is more similar to the input image, the wrong selection may be obtained (since the method is statistical in essence, there may be some false positives or false negatives).

From the analysis above, one can see that there is actually some useful information contained in the projection, but the problem is that the correct projections and the errors introduced by the occluded parts are scalar values and there is no way to separate them.

One interesting observation is that contrary to scalar value, the vector structure has clear separation between elements. Especially in PCA-based vision systems, every element in the vectors corresponds to a pixel value in the images. Since there must be some areas not occluded, there must be some correct pixels in the images. Hence there is a clear separation between the occluded and the non-occluded values in the vector structure. This source of information should be used to create an algorithm which is robust to occlusion. The above analysis leads to a local and global solution which is introduced later. It also leads to a method presented by Leonardis and Bischof [8] where the general framework of traditional PCA-based vision system is employed.

The basic idea of Leonardis and Bischof's [8] work is to avoid using projections in the recognition stage. Instead, they try to find the pixels which belong to the non-occluded part. But given an arbitrary input image, one cannot know which part is occluded. Leonardis and Bischof's solution is to use a random selection algorithm to select a set of points from the input images. Then these points are evaluated to decide whether they belong to the non-occluded parts or to the occluded parts.

The evaluation is based on the assumption that any image of the non-occluded object can be approximated as a linear combination of the principal components. Therefore, the key problem is to determine the coefficients of these principal components. In traditional PCA-based vision systems, these coefficients are generated by projections. As discussed above, for the image of the occluded object, the projections will be contaminated by the errors introduced by occluded parts. In Leonardis and Bischof's method, they formulate this problem as a least squares minimization problem given the randomly selected points. By solving this least squares problem, they can get the coefficients. Then they evaluate these points based on the approximation error. The points with an error below a predefined threshold are assumed to belong to the non-occluded part. If the number of this kind of points is greater than a threshold, they regard this set of points as a hypothesis. To get a robust evaluation of the coefficients, many hypotheses are formed and a selection mechanism is used to select the best.

For the recognition stage, which is the on-line part of a PCA-based vision system, their algorithm incurs a large amount of computation compared with the projection. Indeed this algorithm loses the simplicity of PCA-based vision systems. Further, this method uses a random selection process to select the points used to form the hypotheses. To make sure every possible position is tried, the algorithm must select the points in a systematic way. For example, a technique similar to jittering or stochastic sampling in computer graphics can be used.

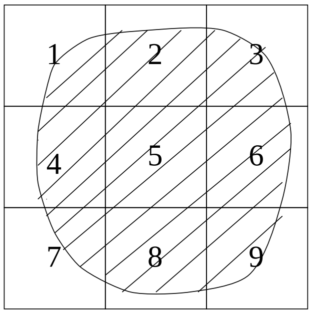
In summary, this method is *ad hoc*. It is not a natural or elegant method although it solves the occlusion problem in some sense.

Pentland, Moghaddam, and Starner introduce a modular method to break down the face into meaningful features such as eigeneyes, and eigennose [12]. This is certainly a local method. However this method is not intended to generate the whole representation of the face. This is fundamentally different from the proposed Mosaic Image Method to be documented later. The features in their method all have clear meanings. In fact, there is a selection stage to select features used. In [6], Krumm does nearly the same thing (using meaningful features) except that he uses an algorithm to select "good" features. These methods are similar to the proposed method in the sense that they are all local methods. However the proposed method is different from their methods in fundamental ways: (1) the ultimate goal of Mosaic Image Method is to generate a representation of the *whole* object; (2) Mosaic Image Method generates the representation of the whole object by slicing the images of the object into many local parts which are called mosaic images; (3) mosaic images can be regarded as features; however, these features usually have no clear meanings such as eyes or noses in Pentland's method, or "good" features in Krumm's work. Instead these mosaic images are just means to generate a global and local representation of the object. The major concern for these mosaic images is not that they have meanings such as eye, nose or "good" features but that the mosaic image size should not be too small (less than the dimension of 8×8 [11]) to contain non-stationary statistics; also they should not be too large to decrease the ability to accommodate occlusion.

It is noteworthy that the ultimate purpose of the proposed method is to generate a global representation just like the representation generated by traditional PCA-based vision systems, yet to generate the global representation in a local way where the local features have no clear meanings (they just serve as the means to generate the global representation and should have proper sizes).

3 Mosaic Image Method (I): Representation

The current PCA-based vision system treats an image as a whole. Consequently, in the covariance matrix, there is an element $COV(x)_{i,j}$ for any two pixels in the image, i.e., to compute the principal



Objects=Features+Topology

Figure 1: *Mosaic Image Method, a local and global method.* One basic rationale for this method is that the pixel in mosaic 1 usually has little correlation with pixels in mosaic 9 or this correlation is relevant.

components it has to account for the correlation of every pair of pixels in the image. This not only makes the computation very expensive, it also makes this method sensitive to occlusion because it depends on the whole image. Furthermore, it is hard to justify that every pixel in the images correlates to each other or that this correlation is relevant.

Alternatively, the images can be treated locally instead of globally. It is the authors' argument that an object is feature based, and that the feature is a local property, which depends only on a small neighbourhood of pixels. The relative positions of these features give the whole structure or topology of the object. Applying this view to PCA, only the correlation of local pixels need to be accounted for and a method called *Mosaic Image Method* is proposed (Fig. 1). In Mosaic Image Method, images are sliced into equal small dimensional mosaic images. Local eigenvectors are generated by accounting for the local correlation in these small mosaic images. Global eigenvectors are formed by these local eigenvectors according to the relative positions of these mosaic images.

Many optical phenomena are local in essence. One of these phenomena is occlusion. As long as some key features are present, the object should be recognizable (Fig. 2). Specular highlight is also a local phenomenon, because it is highly dependent on lighting and view directions. Usually only part of the object has specular highlight and for the most part, the specular highlight can be ignored. Furthermore, cast-shadow can also be treated as occluded parts since cast-shadow usually changes the appearances of objects drastically. Therefore, occlusion, specular highlight and cast-shadow can all be treated in the same way since they share the same properties: (1) all are local phenomena; (2) all change the appearances of parts of objects drastically.

If the image size is $L \times W$, and the size of the small mosaic image is $m \times n$, then the image contains rc mosaic images, where

$$L = mr, W = nc. \tag{2}$$

By raster scanning these mosaic images, vector $\mathbf{v}_{i,j}$ $(i = 0, 1, .., r-1;\ j = 0, 1, .., c-1)$ results. By

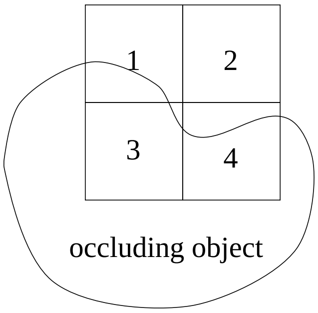


Figure 2: *To recognize an object, at least some key features should be observed.* These key features are usually enough for recognition and reconstruction.

concatenating these vectors, a vector $\mathbf{V}$ for the whole image results:

$$\begin{aligned}\mathbf{V} \;=\; & [\mathbf{v}_{0,0}, \mathbf{v}_{0,1}, ..., \mathbf{v}_{0,c-1}, \\ & \mathbf{v}_{1,0}, \mathbf{v}_{1,1}, ..., \mathbf{v}_{1,c-1}, \\ & ......, ..., \\ & \mathbf{v}_{r-1,0}, \mathbf{v}_{r-1,1}, ..., \mathbf{v}_{r-1,c-1}]. \end{aligned} \tag{3}$$

Applying PCA to each mosaic $\mathbf{v}_{i,j}$, eigenvectors $\Phi_t^{i,j}$ $(t = 0, 1, ..., mn - 1;\ i = 0, 1, ..., r - 1;\ j = 0, 1, ..., c - 1)$ can be computed (these eigenvectors are sorted by non-increasing order). By concatenating these eigenvectors according to the relative positions of corresponding mosaic images, principal components for the whole image result:

$$\Phi_t = [\Phi_t^{0,0}, \Phi_t^{0,1}, ..., \Phi_t^{0,c-1}, ..., \Phi_t^{r-1,0}, \Phi_t^{r-1,1}, ..., \Phi_t^{r-1,c-1}]. \tag{4}$$

This is a new *global* representation of the whole image. Yet the method to generate new eigenimages is via computing local correlations. This is the reason that this method is a local and global method.

The inner product of any two of these principal components is:

$$\Phi_u \cdot \Phi_v = \sum_{i=0,j=0}^{i=r-1,j=c-1} \Phi_u^{i,j} \Phi_v^{i,j} = rc\ \delta_{u,v}, \tag{5}$$

$\delta_{u,v}$ is zero except when $u = v$. Therefore Φ_i $(i = 0, 1, 2, .., min(N - 1, nm - 1))$ (N is the number of the sample images used) are mutually orthogonal. Therefore according to theory of linear algebra, these Φ_i are independent of each other and form part of the basis for the LW dimensional space $\mathbb{R}^{LW}$. However they do not form the complete basis vectors of an LW dimensional space $\mathbb{R}^{LW}$ because $mn < LW$. But again a small number of principal components is enough for the recognition and reconstruction since in essence in every mosaic image the PCA method is applied.

One *interesting and important* thing is about the property of the global representation. According to principal component analysis, the global representation generated by treating an image as a whole is the optimum representation. Therefore although $\Phi_i's$ form part of the basis in the same LW dimensional space $\mathbb{R}^{LW}$, they are certainly not the optimum basis in the context that the image is treated as a whole, i.e, $\Phi_i's$ are used as a whole in the reconstruction and recognition stage. This is verified by experiments presented later. However, this way to apply the new representation Φ_i is certainly not the best way since they are generated locally and globally. Later, experimental results show that using this representation either purely locally or purely globally gives poor performance. However, using this representation both locally and globally give very good performance.

In Mosaic Image Method, Φ_i are applied locally and globally. The input image (after subtracting the mean image from it) is sliced into mosaic images in the same way by which the new representation is generated. The *projection vector* of the input image is therefore:

$$P_t = [p_t^{0,0}, p_t^{0,1}, ..., p_t^{0,c-1}, ..., p_t^{r-1,0}, p_t^{r-1,1}, ..., p_t^{r-1,c-1}], \tag{6}$$

where $t = 0, 1, ..., t$ (t is the number of principal components used).

The projection vectors provide information on the features and the topology information of the objects.

One point that needs to be raised here is that the shape formed by all concerned mosaic images is not necessarily a rectangle. It can be of any shape as long as it encloses all possible shapes of the object. These concerned mosaic images form a mask (Fig. 3).

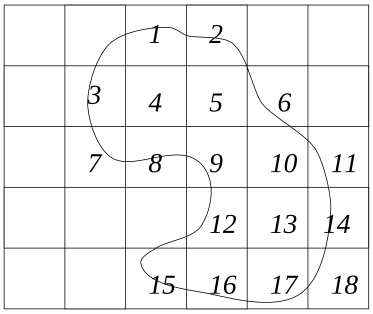


Figure 3: *Mask of an object.* The shape formed by mosaic images can be of any shape as long as the shape encloses all possible shapes of the object. In this example, the irregular mask contains 18 small mosaic images.

4 Experiment to Test the New Representation: A Better Representation than the Optimum Representation

4.1 Experimental Setup

Thorough experiments are conducted to test the new representation generated by Mosaic Image Method. The object used in these experiments is a box with complex texture on it (Fig. 4). The lighting is usually office fluorescent lighting and the background for the image taken is a piece of black flannel. The box is positioned on an accurate turn-table. Every time a new image is to be taken, the table is rotated by 2 degrees first. In all, 180 images of this box are taken by a CCD camera. All these images are in the dimension of 160×120. Ninety images with 4 degree pose difference are used as the sample images.

To apply Mosaic Image Method to compute eigenimages, an image is divided into 20×20 mosaic images (Fig. 4). Therefore there are 48 mosaic images in every image. The mask used in this image

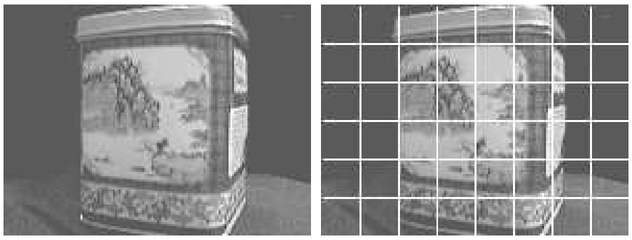


Figure 4: *A box with complex texture.* Every image is divided into 20×20 mosaic images. The image is gamma-corrected for visualization. Here $\gamma = 2$.

is the five columns in the middle. Hence the total mosaic images covered by the mask is 30. The mosaic images in a specific location such as row 4 and column 2 (Fig. 5) are used as the local sample images to compute the local eigenimages. Mosaic Image Method is used to generate the global eigenimages which serve as the new representation. To compare with Mosaic Image Method, the traditional PCA method is also used to compute the eigenimages. The two representations are shown in Fig. 6. The representation by the traditional PCA method is the optimum one. However, to the authors' surprise, it is found that this optimum only holds when the images are treated as a whole. The experimental results to test the representation are presented in the following.

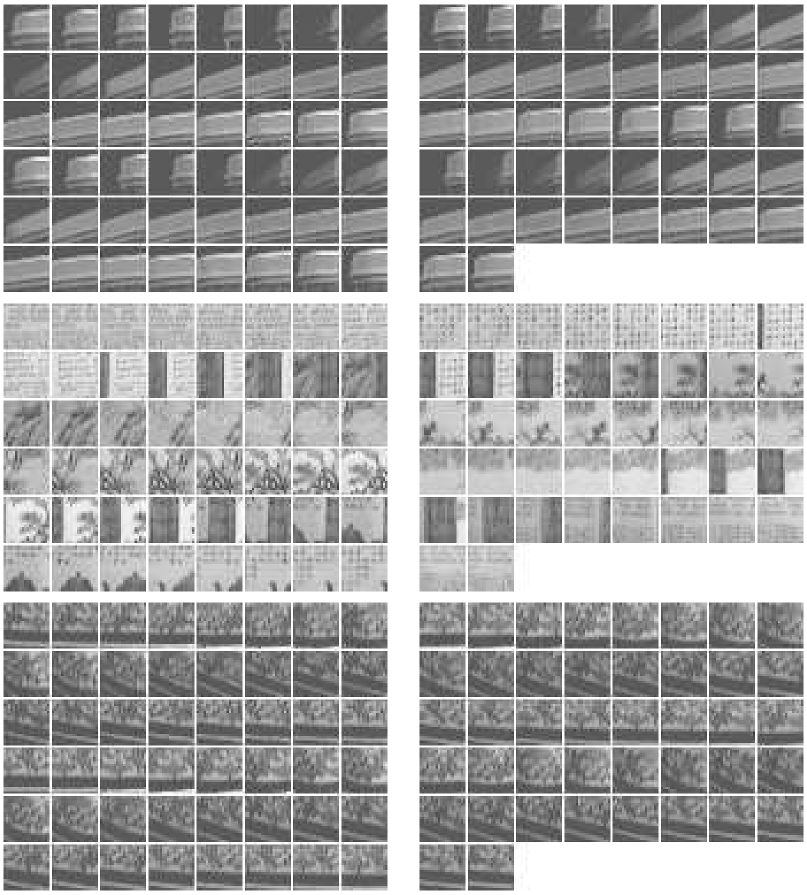
4.2 Using the New Representation Purely Globally

The optimum representation (which forms a basis in the image space) is optimum when the image is treated as a whole. Hence this representation is certainly better than the representation generated by the proposed method if the new representation is used as a whole, i.e., it is used purely globally (although it is generated by a local and global method). By purely globally, it means that new images and eigenimages are all treated as a whole. When an image arrives in the recognition stage, it is not sliced into small mosaic images. When projections are conducted, the eigenimages are not sliced into the small mosaic images either. Instead, the whole image is projected onto the whole eigenimages. The projection scale values instead of projection vectors are generated. In Fig. 7, the new representation is used purely globally in the reconstruction process. It is found that the reconstructed images by the optimum representation are better than the reconstructed images by the new representation where the new representation is used purely globally in the reconstruction process.

4.3 Using the New Representation Locally and Globally by Mosaic Image Method

When the new representation is used by Mosaic Image Method, i.e., used locally and globally, the reconstructed images by the new representation are much better than those by optimum representation. Experimental results confirm this claim (Fig. 8).

In Fig. 8, the first row is the original images. The second and the third rows are reconstructed images by traditional PCA-based method and Mosaic Image Method. The reconstructed images by Mosaic Image Method are much better than those by the traditional PCA-based method. The



(0, 2)

(3, 4)

(5, 3)

The first 48 mosaic images The last 32 mosaic images

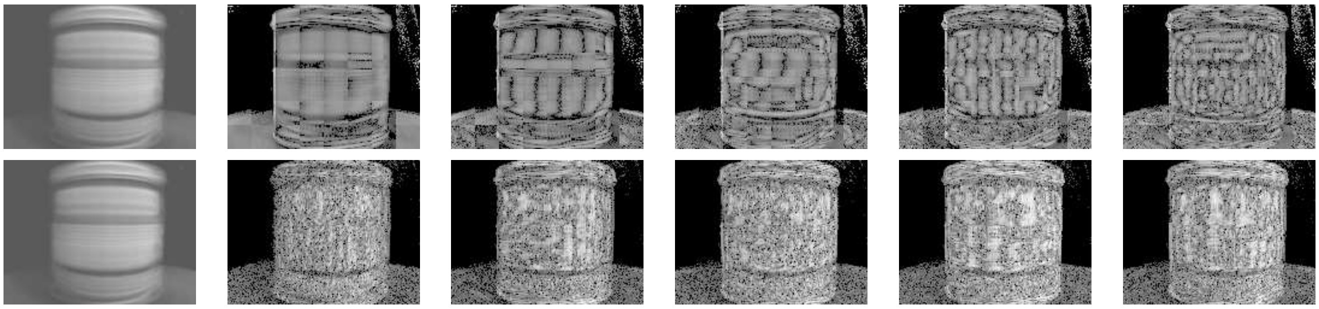
Figure 5: *All sample mosaic images in a special location of the mosaic mask.* Here the location in the mosaic mask are (0, 2), (3, 4) and (5, 3). $\gamma = 2$.

reason for this may be: the traditional representation is generated by accounting for the correlation of every pair pixels in an image. Yet the representation generated by Mosaic Image Method accounts for the local correlation. The global structure is taken care of by the relative positions (topology) of these mosaic images.

One interesting thing is that, until now, very few reconstructed images by the optimum representation have appeared in the literature. The reason, perhaps, is that these reconstructed images are not good.

4.4 Using the New Representation Purely Locally: the New Representation Is A Global Representation

One thing which needs to be emphasized here is that Mosaic Image Method gives a global representation. Certainly this representation can be used purely locally. By purely locally, it means that one may just regard that Mosaic Image Method only gives a simple collection of local PCA representations and ignore internal relationship and global structure of these local PCA representations



mean image eigenimages

Figure 6: *Optimum representation and new representation.* The representations generated by Mosaic Image Method and by traditional method. The first row is the mean image and the first 5 eigenimages by Mosaic Image Method. The second row is the mean image and the first five eigenimages by traditional method where the image is treated as a whole. $\gamma = 5$ for eigenimages, $\gamma = 2$ for mean image.

which give the global information such as topology (it may also include other subtle things). In other words, these local PCA representations are used independent of each other (they do not form a "team").

One direct effect of this purely local point of view is that each mosaic image can be represented by a different number of eigenimages. However it turns out that, in this way, a very poor representation is formed since the reconstructed images are worse than the reconstructed images by treating the representation both locally and globally. In Fig. 9, the reconstructed images by Mosaic Image Method and by purely local method are compared. The first row is the original images. The second and the third rows are reconstructed images by purely local method and Mosaic Image Method. Mosaic Image Method gives a better performance.

In Mosaic Image Method, the number of the eigenimages used, t, is decided by:

$$t = \max_i(t_i), \tag{7}$$

here $i = 0, 1, 2, ..., rc - 1$ (rc is the total number of mosaic images in an image). t_i is the smallest number of eigenimages for mosaic image i which satisfies:

$$\frac{\sum_{i=0}^{i=t_i-1} \lambda_i}{\sum_{i=0}^{i=min(mn,N)-1} \lambda_i} \geq \tau, \tag{8}$$

here $i = 0, 1, 2, ..., rc - 1$. τ ($0 \leq \tau \leq 1$) is a threshold. The mosaic image is $m \times n$, the total number of sample images is N.

However in the purely local method, every mosaic image can use a different number of eigenimages t_i ($i = 0, 1, 2, ..., rc - 1$).

Purely local method is a natural thought when using the new representation. However, just for aesthetic concerns, purely local is not as good as both local and global. Besides, experimental results indicate that this is not a good way to apply the new representation. Furthermore this is not what Mosaic Image Method is originally created for.

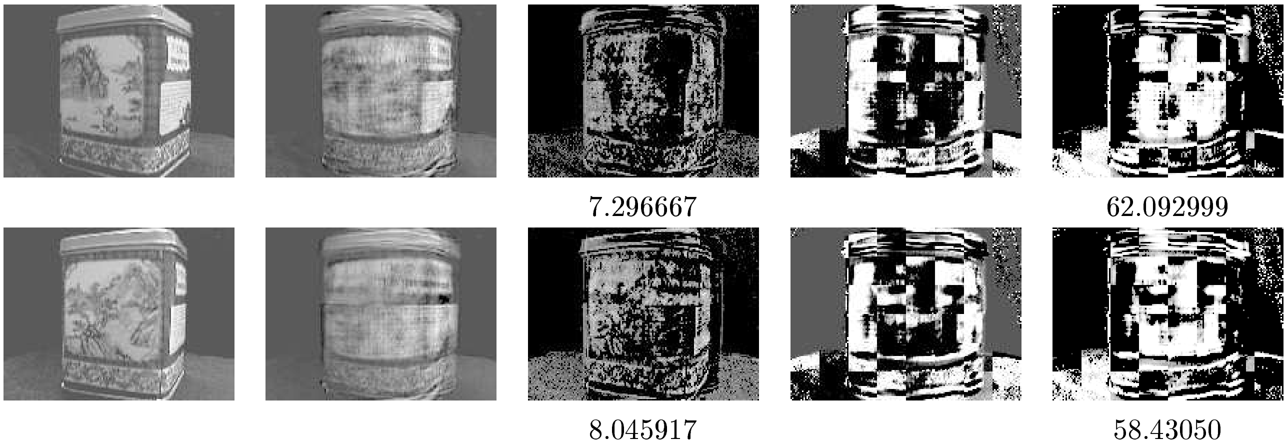


Figure 7: *Using new representation purely globally.* The comparison of the reconstructed images by the optimum representation and new representation while the new representation is used purely globally. The first column is the original images: the image in the first row is a model image while the image in the second row is a new test image. The second and fourth columns are reconstructed images by the optimum representation and the new representation while it is used as a whole. The third and fifth columns are corresponding residual images. The second and fourth rows give the error per pixel of the reconstructed images. 20 eigenimages are used respectively in these two cases. $\gamma = 5$ for the residual images and $\gamma = 2$ for all other images.

5 Mosaic Image Method (II): Simple Recognition and Reconstruction Algorithms

To recognize, a model of the objects must be set up. In the context of PCA-based vision systems, this model includes projections of the sample images in the eigenspace and the eigenimages themselves. In Mosaic Image Method, the model includes the projection vectors and the new representation. A simple recognition algorithm is used to apply this model. This simple algorithm can accommodate up to 53% occlusion and achieves more than 95% correct recognition rate.

In this simple recognition algorithm, if N sample images are used in the training stage and s eigenimages are used in the recognition stage, then there are s projection vectors for every training image. Therefore there are a total of Ns projection vectors in the model. They are denoted as:

$$P_{t,i} = [p_{t,i}^{0,0}, p_{t,i}^{0,1}, ..., p_{t,i}^{0,c-1}, ..., p_{t,i}^{r-1,0}, p_{t,i}^{r-1,1}, ..., p_{t,i}^{r-1,c-1}], \tag{9}$$

where $0 \le t < s, 0 \le i < N$.

When a new image comes, this new image is sliced into the same size (as in the training stage) mosaic images and projected onto the eigenimages by Mosaic Image Method. Hence s projection vectors of the new image result:

$$W_t = [w_t^{0,0}, w_t^{0,1}, ..., w_t^{0,c-1}, ..., w_t^{r-1,0}, w_t^{r-1,1}, ..., w_t^{r-1,c-1}], \tag{10}$$

here $0 \le t < s$.

For every mosaic image in the new input image, the distance from the projections of the new image and the models are computed by Euclidean norm:

$$d_i^{u,v} = \sqrt{\sum_{j=0}^{j=s-1} (w_j^{u,v} - p_{j,i}^{u,v})^2}, \tag{11}$$

here $0 \leq i < N,\ 0 \leq u < r-1,\ 0 \leq v < c-1$. rc is the number of mosaic images in the new input image.

Since there are rc mosaic images in an image, then there are rc $d_i^{u,v}$. To generate the distance of the input image to model image i, the simplest way is by just using one of the rc distances as the whole image's distance to model image i. How to select? At first, the author used the smallest one as the distance. It turns out that this is not good when occlusion occurs. Therefore the author sorted out the distances and observed the performance of all these rc *sorted* distances. As expected, several largest distances do not give good performance; however several smallest distances also do not give good performance. It is found that the best performance is achieved when several middle distances are selected as the distance of the whole image to model image i, i.e.,

$$d_i = \Xi_{u,v}((d_i^{u,v})), \tag{12}$$

here Ξ is the median filter. The explanation for this is as follows. In the recognition stage, the object presented to the vision system is not exactly the learned objects; it may be a totally different object or the same object yet with occluded parts. Due to the very complexity of the texture of the images, there may be some texture which has smaller projection distance yet is not part of the object. In this situation, median filter approximates the usual de-noise operations used in image processing.

Note, here the global distance is generated by integrating the local distances.

After this, the distance of the new image to all model images is determined as:

$$d = \min_i(d_i). \tag{13}$$

The model image which reaches the smallest distance is called the nearest model and denoted as χ. Since the mosaic image which gives the minimum distance can be easily tracked down (denote it as ξ), this mosaic image must be a non-occluded mosaic image. The information in this mosaic image is used to reconstruct the whole image. Denote the projections of this mosaic image as p_i^ξ $(i = 0, 1, ..., s-1)$ and the projections of the corresponding mosaic image in model χ as χ_i^ξ. Then a ratio λ is generated by:

$$\lambda = \Xi_i(\frac{p_i^\xi}{\chi_i^\xi}), \tag{14}$$

here Ξ is the median filter. If $\chi_i^\xi = 0$, assume:

$$\frac{p_i^\xi}{\chi_i^\xi} = 1. \tag{15}$$

Then the projections of every mosaic image are computed as:

$$p_i^t = \lambda \chi_i^{\xi}, \tag{16}$$

here $0 \leq t < s,\ 0 \leq i < s - 1$.

The recognition and reconstruction algorithms are simple. However the recognition algorithm already achieves a high correct recognition rate. The reconstruction algorithm also gives good results. Sophisticated recognition algorithms such as neural nets and more sophisticated reconstruction algorithms are certainly worth trying and are directions for further research.

6 Experimental Results

To test the capacity of Mosaic Image Method to accommodate occlusion, different degrees of occlusion are introduced to the test images. Thorough experiments are conducted using more than 110,000 different occlusion images. In this section the experimental results of the recognition and reconstruction algorithms are presented.

6.1 Generating Test Image with Occluded Parts

Test images with different occluded parts are generated by two different methods.

In the first method, some solid squares are introduced to the test images. These solid squares are in the same dimensions of the mosaic images. They can form a big rectangle or can spread in a random way. Further, every solid square occluder can cover exactly one mosaic image; it can also cross mosaic image borders (hence it covers more than one mosaic image). The intensity of the solid square can be any value from 0 to 255. In Fig. 10, some test images generated by this method are given.

In the second method, the occluded parts are parts of a candy box (Fig. 11) which has complex texture. These occluded parts are added to the test images in the same way as the first method. Some images are shown in Fig. 12.

6.2 All Test Images Without Occlusion are Correctly Recognized

As described in setting up the experiment, in the total 180 images, there are 90 test images. These images are images without occlusion. All these 90 test sample images are correctly recognized by the proposed simple recognition algorithm. Hence the recognition rate for the non-occluded images is 100%.

6.3 Experiment 1: Occlusion by Solid Squares

This experiment consists two parts. In the first part, different sizes of rectangular occluded part are introduced to the 90 test images. Every possible location to position this solid rectangle is tested (in all except one experiment, every square occluder covers only one mosaic image). For example, for a 2×2 rectangle, there are in all 20 positions in the mask of the object where the solid rectangle can be placed. Every solid rectangle is tested for 14 different intensities, i.e., the intensities of the

Table 1: *The results of the experiments with rectangular solid occluded parts.*

occluder size	number of test images	number of incorrect recognition	correct recognition rate
1×1	37800	0	100%
2×2	25200	0	100%
3×3	15120	0	100%
4×3	11340	2	99.98%
3×4	10080	1	99.98%
3×5	5040	36	99.28%
5×3	7560	11	99.85%
3×3 (*move*)	7560	39	99.5%
4×4	7560	122	98.4%
5×4	5040	3830	24.01%
4×5	3780	2645	30.03%

Table 2: *The results of the experiments with the occluder squares in random distribution.*

template	number of square occluders	number of test images	number of incorrect recognition	correct recognition rate
1	12	1260	0	100%
2	15	1260	0	100%
3	16	1260	22	98.25%
4	16	1260	16	99.73%

solid rectangle can be 0, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 240, or 255. In the second part of the experiment, the solid square occluders are introduced in random distribution with total square occluder numbers of 12, 15, 16 and 16 respectively (Fig. 13). The intensity of these square occluders can also be any of the 14 intensities mentioned above. Some test images are shown in Fig. 13. The results of the first part of this experiment are summarized in Table 1. (Note, in the eighth row of Table 1, "move" means one small square occluder can cover more than one mosaic image.) The results of the second part of this experiment are summarized in Table 2.

6.4 Experiment 2: Occlusion by Real Objects

In this experiment, the only difference is that the solid square occluder is replaced by part of a candy box. Different sizes of occluded parts are introduced to the 90 test images to form new test images. Every possible location to put the occluded part is tested. Besides the rectangular occluded part, the occluders are also introduced in random distribution. Four different templates are used (Fig. 14). The recognition algorithm reaches 95.47% correct recognition rate in the presence of more than 53% occluded area. The results are shown in Table 3 and Table 4. (Note, in the eighth row of Table 3, "move" means one small square occluder can cover more than one mosaic image.) Some test images with different occluded parts and the reconstructed images after recognition are shown in Fig. 14.

Table 3: *The results of the experiment with the rectangular occluded part of a candy box.*

occluder size	number of test images	number of incorrect recognition	correct recognition rate
1×1	2700	0	100%
2×2	1800	0	100%
3×3	1080	0	100%
4×3	810	0	100%
3×4	720	0	100%
3×5	360	8	97.78%
5×3	540	12	97.78%
3×3 $(move)$	540	0	100%
4×4	540	25	95.47%
5×4	360	201	34.17%
4×5	270	214	26.7%

Table 4: *The results of the experiment with distributed occluders of a candy box.*

template	number of square occluders	number of test images	number of incorrect recognition	correct recognition rate
1	12	90	0	100%
2	15	90	2	97.78%
3	16	90	3	96.67%
4	16	90	3	96.67%

7 Summary

In this paper, Mosaic Image Method is presented. This method can generate a better representation and is very robust to occlusion. The authors would like to emphasize that 53% occluded area is a very high occlusion rate (see Fig. 13, Fig. 14). In such a high occlusion rate, even for humans, it is hard to tell what the image is (see Fig. 13, Fig. 14). In the following the main points of this paper are summarized.

1. The representation generated by Mosaic Image Method is a better representation than the optimum representation by the traditional PCA method. The reason for this is that the representation is a local and global representation. It accounts for local correlation by local PCA method and global information by the relative relations of these mosaic images. The reconstructed images by the new representation generated by the proposed method are much better than those by traditional PCA method.

2. Mosaic images in Mosaic Image Method have no clear meanings such as eyes, or noses of eigenfeatures, or "good" features. Instead the major concern for a mosaic image is its size. The size should not be too small to contain non-stationary statistics or too big to decrease its ability to accommodate occlusion.

3. The proposed representation is a local and global representation. Using this new representation locally and globally is also the way that this representation is applied for reconstruction and recognition.

4. By using a simple recognition algorithm, Mosaic Image Method can accommodate up to 53% occlusion with more than 95% correct recognition rate. To the author's knowledge, this is the best result in the presence of occlusion in PCA-based vision systems.

5. The simple reconstruction method gives very good reconstructed images.

One interesting research direction is to apply more sophisticated recognition and reconstruction algorithms since there is much information contained in the projection vectors and residuals of the mosaic images. Preliminary investigation suggests that this structure conforms to the structures of neural nets. However, the 2D structure of projection vectors and residual information are quite complex. How to combine this information with the neural net structures is an interesting and subtle problem.

References

[1] Belhumeur, P., Hespanha, J., and Kriegman., D., Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *European Conference on Computer Vision*, pp. 45-48, April 1996.

[2] Cui, Y., Swets, D., and Weng, J., Learning-Based Hand Sign Recognition Using SHOSLIF-M, *Proceedings of International Conference on Computer Vision,* pp. 631-636, Cambridge, Massachusetts, June 1995.

[3] Foley, J.D., van Dam, A., Feiner, S.K., and Hughes, J.F., *Computer Graphics: Principles and Practice*, Addison-Wesley Press, 1992.

[4] Horn, B.K., *Robot Vision*, MIT Press, 1986.

[5] Kohonen, T., Riittinen, H., Jalanko, M., and Haltsonen, S., A Thousand-Word Recognition System Based on the Learning Subspace Method and Redundant Hash Addressing, *International Conference on Pattern Recognition*, pp. 158-165, Palm Beach, Florida, 1980.

[6] Krumm, J., Eigenfeatures for Planar Pose Measurement of Partially Occluded Objects *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 55-60, San Francisco, California, July 1996.

[7] Kukunaga, K., *Introduction to Statistical Pattern Recognition*, 2nd Ed., Academic Press, 1990.

[8] Leonardis, A., and Bischof, H., Dealing with Occlusion in the Eigenspace Approach, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 270-277, San Francisco, California, July 1996.

[9] Moghaddam, B., and Pentland, A., Probabilistic Visual Learning for Object Recognition, *Proceedings of International Conference on Computer Vision*, pp. 786-793, Cambridge, Massachusetts, June 1995.

[10] Murase, H., and Nayar, S., Learning and Recognition of 3D Object from Appearance, *Proceedings of IEEE Workshop on Qualitative Vision*, pp. 39-50, June 1993.

[11] Netravali, A.N., and Haskell, B. G., *Digital Pictures*, Plenum Press, 1988.

[12] Pentland, A., Moghaddam, B., and Starner, T., View-Based and Modular Eigenspace for Face Recognition, *CVPR '94*, pp. 84-91, Seattle, June, 1994.

[13] Turk, M., and Pentland, A., Eigenface for Recognition, *Journal of Cognitive Neuroscience*, Vol. 3(1), pp. 71-96, 1991.

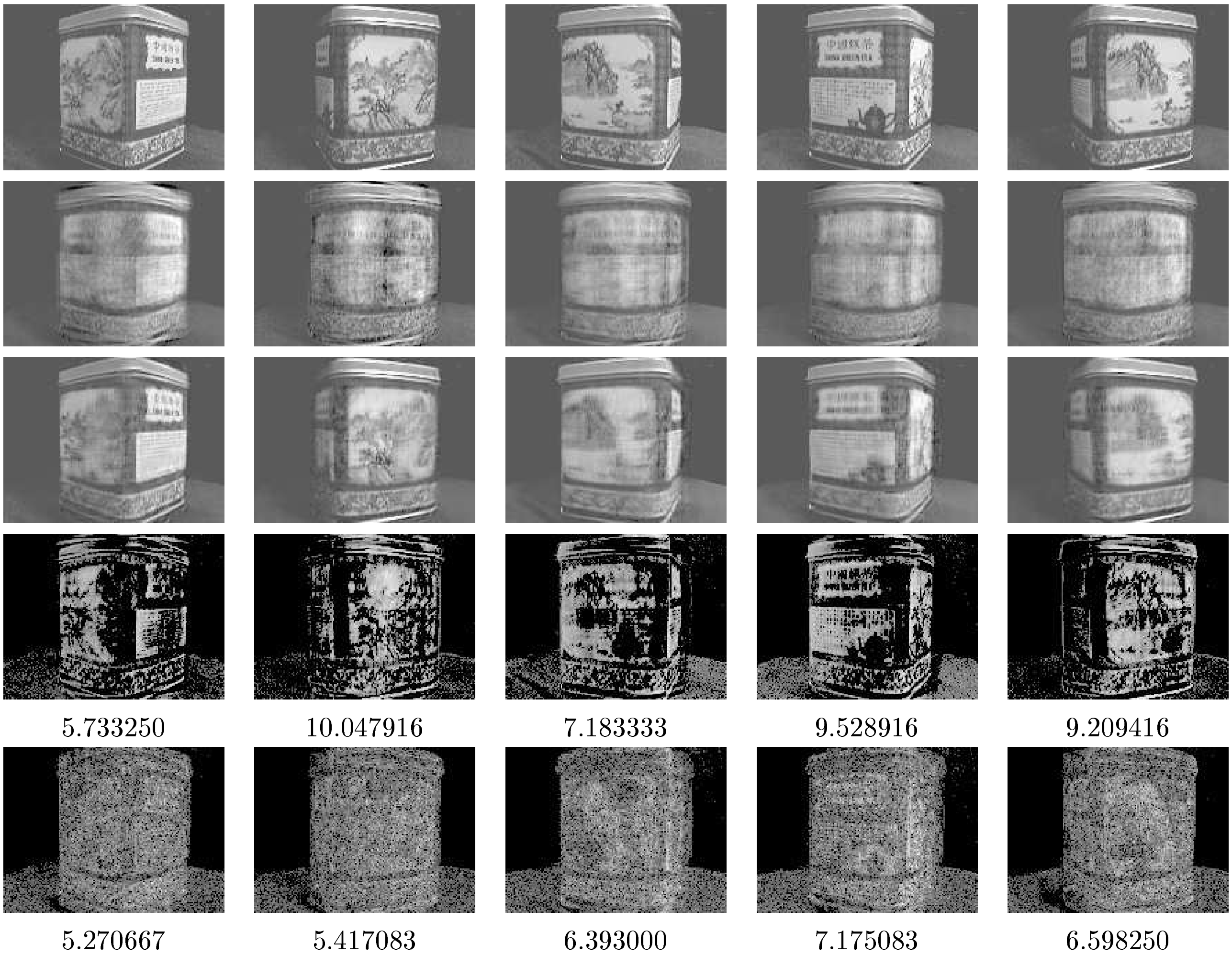


Figure 8: *Using new representation locally and globally by Mosaic Image Method.* The comparison of reconstructed images by the optimum representation and the new representation. The first row is the original images in which the first two are sample images while the other three images are new test images. The second and third row are reconstructed images by optimum representation and by new representation. The fourth and sixth rows are corresponding residual images. The fifth and seventh rows are error per pixel for the reconstructed images. 20 eigenimages are used respectively in the two methods. $\gamma = 5$ for residual images and $\gamma = 2$ for all other images.

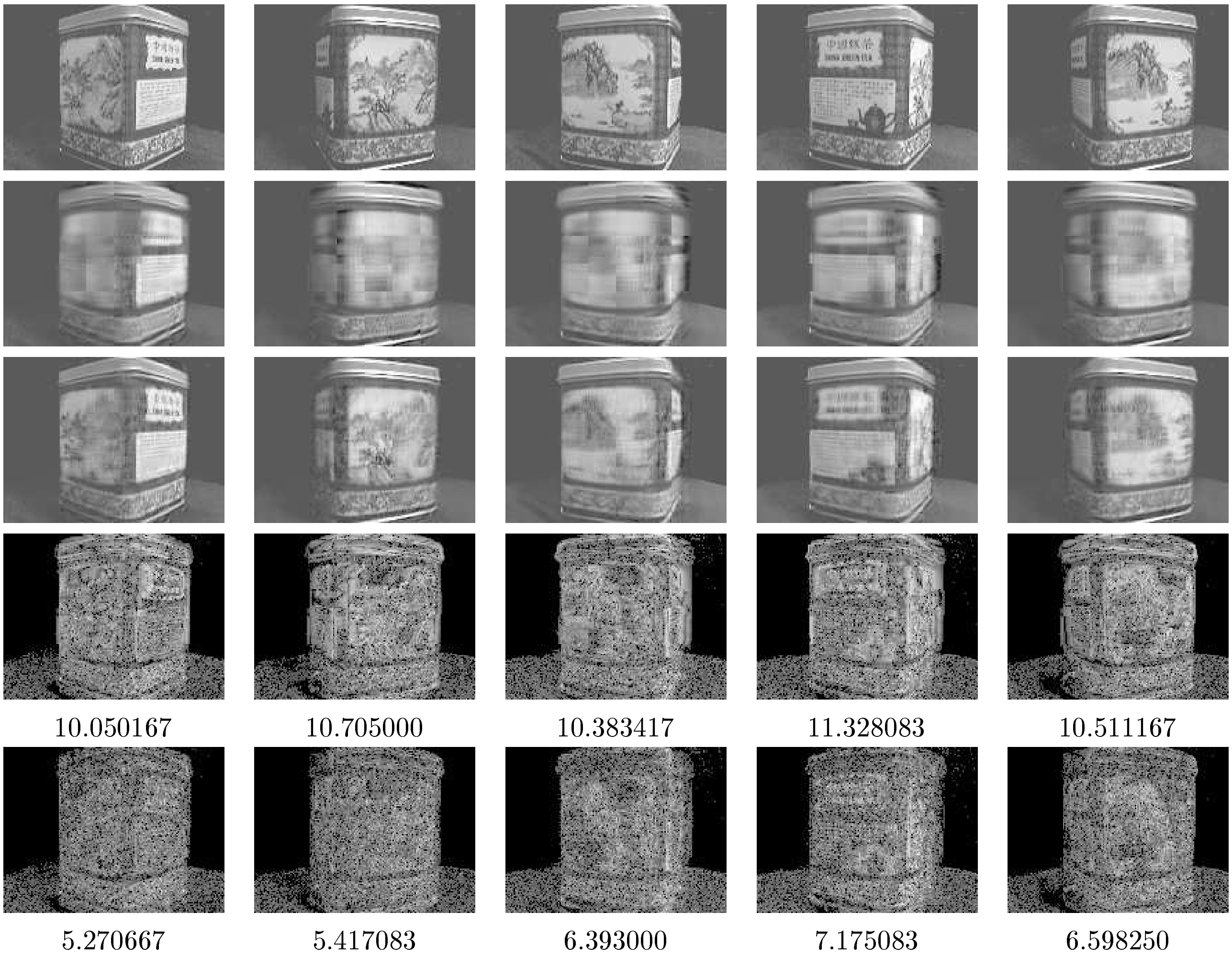


Figure 9: *Using new representation purely locally.* The comparison of reconstructed images by Mosaic Image Method method and purely local method. The first row is the original images in which the first two are sample images and the other three images are new test images. The second and third rows are reconstructed images by purely local method and by Mosaic Image Method. The fourth and sixth rows are corresponding residual images. The fifth and seventh rows are error per pixel for the reconstructed images. $\tau = 0.7$ in both Mosaic Image Method and purely local method. 20 eigenimages are used in Mosaic Image Method. Different numbers of eigenimages for different mosaic images are used in the purely local method. $\gamma = 5$ for residual images and $\gamma = 2$ for all other images.

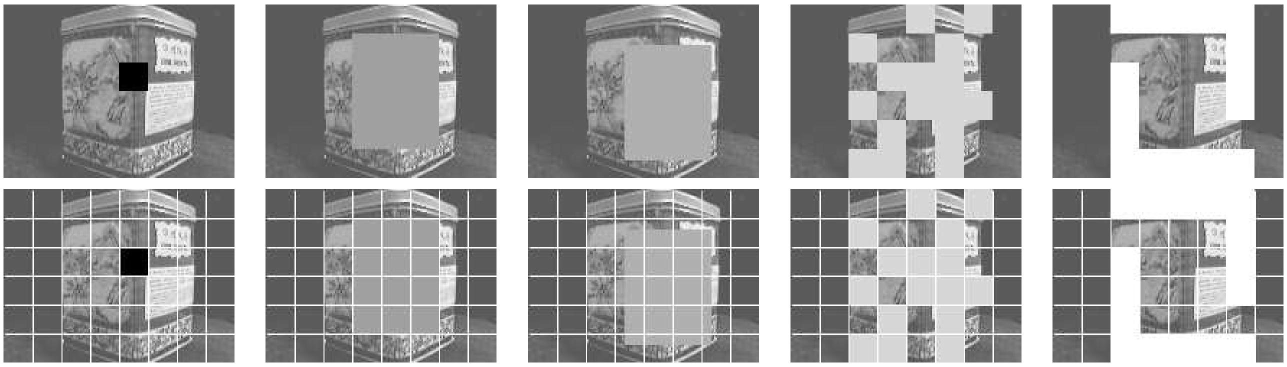


Figure 10: *Test images with solid square occluders.* The images in the first and second rows are the same except that in the second row images, some lines are added. The intensities of the solid occluders are 0, 100, 120, 180, 255 respectively. In the third column, every solid square occluder covers more than one mosaic image. Therefore, there are actually 20 mosaic images affected in this image, not just 12 mosaic images affected. The solid squares can form a rectangle or be randomly distributed. $\gamma = 2$ in all these images.

Figure 11: *A candy box with complex texture.* $\gamma = 2$.

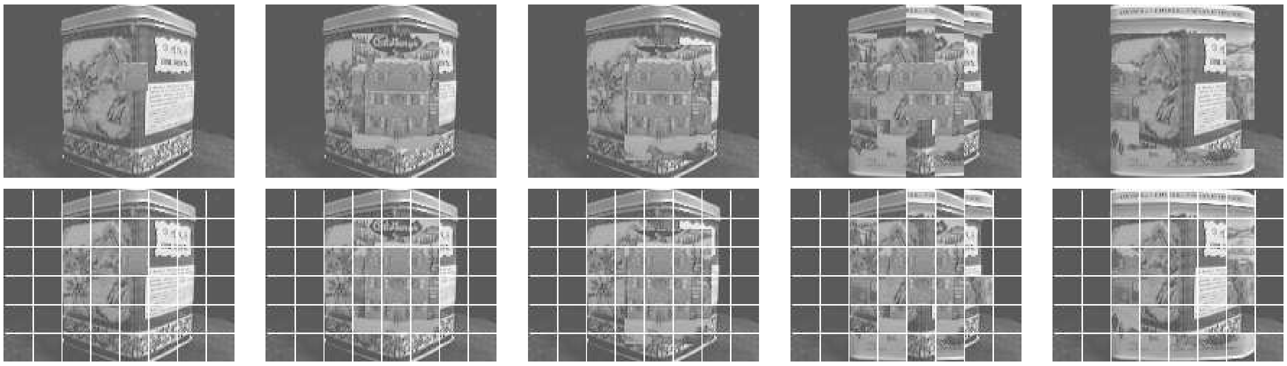
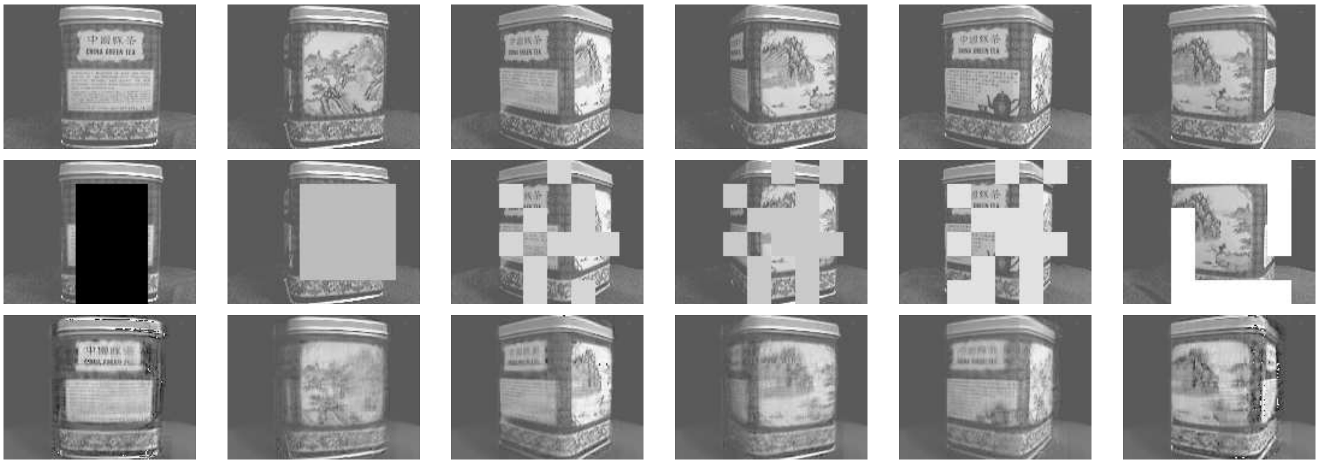


Figure 12: *Test images with parts of a candy box as occluders.* The images in the first and second rows are the same except that in the second row images, some lines are added. In the third column, every square occluder covers more than one mosaic image. Therefore, there are actually 20 mosaic images affected in this image, not just 12 mosaic images affected. The square occluders can form a rectangle or be randomly distributed. $\gamma = 2$ in all these images.



5×3 | 4×4 | template 1 | template 2 | template 3 | template 4

Figure 13: *Some experimental results with solid square occluders.* The first row is the original test images. The second row is the test images with occluded parts. The third row is the reconstructed images after recognition. $\gamma = 2$.

5×3 | 4×4 | template 1 | template 2 | template 3 | template 4

Figure 14: *Some experimental results with occluded parts of a candy box.* The first row is the original test images. The second row is the test images with different occluded parts. The third row is the reconstructed images by the simple reconstruction algorithm after recognition. $\gamma = 2$ in all these images.