6

Occluded Image Object Recognition using Localized Nonnegative Matrix Factorization Methods

Ivan Bajla[1], Daniel Soukup[2] and Svorad Štolc[3]
[1,2]*Austrian Institute of Technology GmbH, Seibersdorf*
[3]*Institute of Measurement Science, Slovak Academy of Sciences, Bratislava*
[1,2]*Austria*
[3]*Slovakia*

1. Introduction

In the nineties, appearance-based methods for image object detection/recognition have evoked a renewed attention in computer vision community thanks to their capability to deal with combined effects of shape, illumination conditions, and reflectance properties in the scene (Beymer & Poggio, 1995; Mel, 1997; Murase & Nayar, 1995; Turk & Pentland, 1991; Yoshimura & Kanade, 1994). The major advantage of these methods is that both learning and recognition stage of image processing utilize only two-dimensional brightness images without an intermediate processing. On the other hand, the most severe limitation of these approaches (in their conventional form) consists in problems with object occlusions and varying background. The basic characteristic of the appearance-based approaches is as follows.

They consist of the two stages: the off-line training (learning) stage and the on-line recognition stage. In the first stage a set of sample images (templates) are available which encompass the appearance of a single object under various conditions (Yoshimura & Kanade, 1994), or multiple instances of a class of objects, e.g., faces (Turk & Pentland, 1991). The images in sample sets are chosen to be correlated, thus enabling efficient compression using Principal Component Analysis (PCA) (Jolliffe, 2002). In the second recognition stage, given an unknown input image, we project this image (of identical size as the training images) to the eigenspace generated in the first stage. The recovered coefficients indicate the particular instance of a class to which the given input image belongs. This process can equally be applied to sample objects and subimages of an image in which the existence and/or position of a template object should be detected.

Leonardis & Bischof (1994) modified the PCA space representation method with the goal to improve recognition rates for cases with occlusions. Their robust method extended the domain of applicability of the appearance-based methods towards more complex scenes which contain occlusions and background clutter. The basic novelty of their approach consists in the way the coefficients of the eigenimages are calculated. Instead of computing them by a standard projection of the input data onto an eigenspace, they calculate the coefficients of linear combinations of eigenimages using an objective function and hypotheses on object point subsets. Indeed, this method provides a reduction of occlusion problems. However,

this improvement is reached at the expense of significant increase of the computational cost of operations exactly in the on-line stage of the PCA method. The method requires tuning eight specific parameters and a number of additional procedures to be implemented within the on-line recognition stage, thereby reducing the main advantage of PCA data representation for on-line recognition applications: a simple projection on an eigenspace and the nearest neighbor search.

Regardless of the weak points of PCA and its more robust modifications, subspace data representation methods are still a challenging branch of object recognition methods used in computer vision and pattern recognition. In particular, these methods find applications in the fields of face identification, recognition of digits and characters occurring at various labeled products. Therefore we were interested in exploration of other possibilities of object recognition that would be robust to occlusions and could parallely provide an acceptable solution for applications requiring high-performance on-line image processing.

Lee & Seung (1999) showed for the first time that for a collection of face images an approximative representation by basis vectors, encoding the mouth, nose, and eyes, can be obtained using a Nonnegative Matrix Factorization (NMF). It is a method for generating a linear representation of data using nonnegativity constraints on the basis vector components and encoding coefficients. The nonnegative matrix decomposition can formally be described as follows:

$$\boldsymbol{V} \approx \boldsymbol{W} \cdot \boldsymbol{H}\,, \tag{1}$$

where $\boldsymbol{V} \in \mathcal{R}^{n \times m}$ is a positive image data matrix with n pixels and m image samples (templates, which are usually represented in lexicographic order of pixels as column-vectors), $\boldsymbol{W} \in \mathcal{R}^{n \times r}$ are basis column vectors of an NMF-subspace, and $\boldsymbol{H} \in \mathcal{R}^{r \times m}$ contains coefficients of the linear combinations of the basis vectors needed for reconstruction of the original data (called also encoding vectors). Usually, r is chosen by the user so that $(n+m)r < nm$. Then each column of the matrix $\boldsymbol{W}$ represents a basis vector of the generated NMF-subspace. Each column of $\boldsymbol{H}$ represents the weights needed to linearly approximate the corresponding column in $\boldsymbol{V}$ (image template) by means of the vector basis $\boldsymbol{W}$. Various error (cost) functions were proposed for NMF (Lee & Seung, 2001; Paatero & Taper, 1994). The most frequently used is the Euclidean distance:

$$E(\boldsymbol{W}, \boldsymbol{H}) = \|\boldsymbol{V} - \boldsymbol{W} \cdot \boldsymbol{H}\|^2 = \sum_{i,j} (V_{i,j} - (\boldsymbol{WH})_{ij})^2\,. \tag{2}$$

The main difference between NMF and other classical factorization models relies in the nonnegativity constraints imposed on both the basis vectors of $\boldsymbol{W}$ and encoding vectors of $\boldsymbol{H}$. In this way, only additive combinations are possible:

$$(\boldsymbol{V})_{i\mu} \approx (\boldsymbol{WH})_{i\mu} = \sum_{j=1}^{r} W_{ij} H_{j\mu}\,.$$

Increasing interest in this factorization technique is due to the intuitive nature of the method that provides extraction of additive parts of data sets interpretable as real image parts, while reducing the dimensionality of the input data at the same time. In the recent years several modifications of NMF schemes applied to various types of image data have been proposed and explored. Also mathematical issues of optimization of objective functions defined for NMF have been addressed and improved numerical algorithms have been developed. We only mention the best-known of them:

1. Local Nonnegative Matrix Factorization (Feng et al., 2002)
2. Nonnegative Matrix Factorization (Liu et al., 2003)
3. Nonnegative Sparse Coding (Hoyer, 2002)
4. Nonnegative Matrix Factorization with Sparseness Constraints (Hoyer, 2004)
5. Discriminant Nonnegative Matrix Factorization (Buciu, 2007; Buciu et al., 2006; Buciu & Pitas, 2004)
6. Nonsmooth Nonnegative Matrix Factorization (Pascual-Montano et al., 2006)
7. Learning Sparse Representations by Nonnegative Matrix Factorization and Sequential Cone Programming (Heiler & Schnörr, 2006).

In our previous research we focused on studying the influence of the matrix sparseness parameter on recognition rates, in particular in images with occluded objects (Bajla & Soukup, 2007). In the recognition experiments, carried out with this goal, we also studied four types of metrics used in the nearest neighbor search. We proposed a weaker alternative of NMF (the so-called semi-NMF) based on Hoyer's NMF algorithm (Soukup & Bajla, 2008) that is numerically more stable.

2. Parts-Based methodology of NMF

In the seminal paper Lee & Seung (1999), the methodology of nonnegative matrix factorization was applied for the first time to the task of image representation. Lee and Seung motivated their approach by psychological and physiological evidence for *parts-based* representation in the brain, and by certain computational theories. However, the notion of *parts-based* representation was not introduced as a formal term. They stated that the NMF algorithm is able to learn parts of the face images and the core of this ability stems from the nonnegativity constraints included in NMF. They also compared the proposed NMF basis vectors to conventional PCA bases with holistic structure and claimed that NMF bases better correspond with intuitive notion of the parts of a face. Moreover, they argued that "PCA allows complex cancellation between positive and negative terms in the linear combination of basis vectors (eigenimages) and therefore it lacks the intuitive meaning of adding parts to form a whole". In the paper of Lee and Seung, an illustration is given of the NMF basis face images (matrix $\boldsymbol{W}$) and face image encodings (matrix $\boldsymbol{H}$). The sparseness of basis images is explained by their non-global nature (they contain several versions of mouths, noses, eyes, etc.), while the sparseness of the encoding coefficient matrix is attributed to the ability of the method to include some basis images and to cancel others from the linear combinations given by the product $\boldsymbol{W} \cdot \boldsymbol{H}$. On the basis of the Lee and Seung methodological statements related to the intuitive notion of the parts-based representation of images (in particular, faces), we can summarize that some characteristic regions of the input image, occurring in certain geometrical locations, are understood as image parts which are represented by image basis vectors (columns of the matrix $\boldsymbol{W}$) only in indirect way.

The results of Lee and Seung encouraged researchers to apply the NMF approach to various image object recognition problems, especially to those affected by local deformations and partial occlusions. In the papers Hoyer (2004); Kim et al. (2005); Li et al. (2001); Pascual-Montano et al. (2006), the use of NMF in recognition tasks has been further explored. In Li et al. (2001) the concept of NMF that non-subtractive combining of NMF basis vectors results in forming the whole (image) was confirmed to some extent. However, the authors showed that additive parts learned by NMF are not necessarily localized. On the basis of recognition experiments they also showed that original NMF representation

yields low recognition rates for occluded images. Thus, the results of Li et al. made the justification of the parts-based principle of NMF and its use for the object recognition task questionable. Although they proposed an improved modification of NMF (the so-called local NMF (LNMF)), for learning a spatially more localized parts-based image representation, they did not perform a sufficient number of recognition experiments which could prove better performance of the LNMF method in practical tasks of object recognition.

Buciu & Pitas (2004) developed a novel Discriminant NMF (DNMF) algorithm by introducing two additional constraints on the coefficients. The first constraint is based on the within-class scatter matrix of the class samples (input images) around their mean. The second constraint reflects the between-class variance and it is given by the scatter of the class mean around the global mean. The constraints were incorporated into the divergence cost function of NMF that was applied to the problem of recognizing six basic facial expressions from face images of Kanade et al. (2000) AU-coded facial expression database. The influence of partial occlusions on recognition rates has not been explored systematically neither in this paper, nor in the paper of Li et al.

Kim et al. (2005) explored efficient image representation using Independent Component Analysis (ICA) in the task of face recognition robust to local distortion and partial occlusions. They included in the research also the LNMF method and proved that additional constraints of Li et al. (2001) involved into this method only focused on locality and they do not guarantee localization of meaningful facial features in their basis images.

The next attempt to assign a more accurate meaning to the parts-based methodology of the NMF subspace representation was made in Hoyer (2002; 2004). Hoyer pointed at the most useful property of NMF that is generation of a sparse representation of the data. He stated that such a representation encodes much of the data using few "active" components which make the encoding easy to interpret. Hoyer also claimed that sparseness of basis vectors and encoding coefficients of NMF is reached as a side effect rather than a goal. He proposed a novel NMF modification in which the sparseness of the column vectors in the matrix $\boldsymbol{W}$, as well as the sparseness of the column vectors of the matrix $\boldsymbol{H}$ are explicitly controlled in the course of optimization of the objective function.

We recall here the concept of the vector or matrix sparseness and its measure as was used in Hoyer (2004). The concept of the sparse encoding refers to the data representation task in which only several units are efficiently used to represent typical data vectors. In practice this implies most entries having values close to zero while only few take significantly non-zero values. Various sparseness measures have been used in the literature as mappings from $\mathcal{R}^n \rightarrow \mathcal{R}$, quantifying the amount of energy of a vector packed into a few components. On a normalized scale, the sparsest vector with a single non-zero component should have the sparseness measure equal to one, whereas a vector with no element equal to zero should have a sparseness of zero.

Applying the concept of the sparseness to the NMF task leads to the basic question: what actually should be sparse? The basis vectors of $\boldsymbol{W}$ or the encoding coefficients represented by the matrix $\boldsymbol{H}$? According to Hoyer's claim, such a question cannot be answered in a general way, it all depends on the specific application. E.g., when trying to learn useful features from a database of images, it makes sense to require both $\boldsymbol{W}$ and $\boldsymbol{H}$ to be sparse, signifying that any given object is present in a few images and affects only a small part of the image. Hoyer (2004) derived a projected gradient descent algorithm for NMF with sparseness (the details are given in his paper). It can be briefly described in the following way.

Given any vector $\mathbf{x}$, find the closest (in the Euclidean sense) nonnegative vector s with a given L_1 norm and a given L_2 norm. We start by projecting the given vector onto the hyperplane $\sum s_i = L_1$ by assigning $s_i := x_i + (L_1 - \sum x_i)/\dim(\mathbf{x}), \forall i$. Next, within this space, we project to

the closest point on the joint constraint hypersphere. This is done by moving radially outward from the center of the sphere (the center is given by the point where all components have equal values). If the result is completely nonnegative, we have arrived at our solution. If not, those components that attained negative values must be fixed zero, and a new point is found in a similar fashion under those additional constraints.
Thus, using the sparseness concept, the following modified NMF problem can be formulated in which the sparseness of the factor matrices $\boldsymbol{W}$ and $\boldsymbol{H}$ is explicitly controlled during the optimization process. Given a nonnegative data matrix $\boldsymbol{V}$ of size $n \times m$, find the nonnegative matrices $\boldsymbol{W}$ and $\boldsymbol{H}$ of sizes $n \times r$ and $r \times m$, respectively, such that

$$E(\boldsymbol{W}, \boldsymbol{H}) = \|\boldsymbol{V} - \boldsymbol{W} \cdot \boldsymbol{H}\|^2 \tag{3}$$

is minimized, under optional constraints

$$s(\mathbf{w}_i) = s_W \ , \ \ \forall_i, \ i = 1, \cdots, r,$$
$$s(\mathbf{h}_i) = s_H \ , \ \ \forall_i, \ i = 1, \cdots, r,$$

where $\mathbf{w}_i$ is the i-th column of $\boldsymbol{W}$, $\mathbf{h}_i$ is the ith row of $\boldsymbol{H}$. Here r denotes the dimensionality of an NMF subspace spanned by the column vectors of the matrix $\boldsymbol{W}$, and s_W and s_H are their desired sparseness values. The sparseness criteria proposed in Hoyer (2004) use a measure based on the relationship between L_1 and L_2 norm of the given vectors $\mathbf{w}_i$ or $\mathbf{h}_i$. In general, for the given n-dimensional vector $\mathbf{x}$ its sparseness measure $s(\mathbf{x})$ is defined by the formula:

$$s(\mathbf{x}) = \frac{\sqrt{n} - \|\mathbf{x}\|_1 / \|\mathbf{x}\|_2}{\sqrt{n} - 1} \ . \tag{4}$$

This measure quantifies how much energy of the vector is packed into a few components. This function evaluates to 1 if and only if the given vector contains a single non-zero component. Its value is 0 if and only if all components are equal. It should be noted that the vector scales $\boldsymbol{w}_i$ or $\boldsymbol{h}_i$ have not been constrained yet. However, since $\boldsymbol{w}_i \cdot \boldsymbol{h}_i = (\boldsymbol{w}_i \lambda) \cdot (\boldsymbol{h}_i / \lambda)$, we are free to arbitrarily fix any norm of either one. In Hoyer's algorithm the L_2 norm of $\boldsymbol{h}_i$ is fixed to unity. In this study we have re-run computer experiments with Hoyer's NMF method. The computer experiments, including partially occluded face parts, yielded recognition rates similar to the previously published versions of NMF. In spite of the advantages of the explicit sparseness control in the NMF optimization algorithm proposed by Hoyer, we do not see any direct relation of the sparseness to image parts, as we intuitively understand them. Although it ensures basis vectors with many zeros and local non-zero components, these have not any clear relation to locally defined image parts (regions).
Recently, Spratling (2006) investigated how NMF performs in realistic environments where occlusions take place. As a basic benchmark task he chose a bars problem that consists of a system of elementary bar patterns. He tested NMF algorithms also on the face images from the CBCL and ORL face databases. Based on the results obtained in a comprehensive set of comparative computer experiments he claimed that NMF algorithms can, in certain circumstances, learn localized image components, some of which appear to roughly correspond to parts of the face, but others of which are arbitrary, but localized blobs. According to Spratling, the NMF algorithms essentially select a subset of the pixels which are simultaneously active across multiple images to be represented by a single basis vector. In the case of faces, large subsets of the training images contain virtually identical patterns of the pixel values (eyes, nose, mouth, etc.). The NMF algorithms form distinct basis vectors to represent pieces of these recurring patterns. Spratling concludes that the separate

representation of sub-patterns is due to constraints imposed by the algorithms and is not based on evidence contained in the training images. Hence, while these constraints make it appear that NMF algorithms have learned face parts, these algorithms are representing arbitrary parts of larger image features.
Summarizing the above mentioned results and statements of several authors on the NMF representation of occluded images, we claim that

- the notion of "local representation" is not identical to the notion of "parts-based representation",
- the "sparse representation" does not automatically yield "parts-based representation", and
- the "parts-based representation" of images, as proposed in the published papers, provides no guarantee of achieving satisfactory recognition rates in cases with object occlusions.

3. A modification of the NMF for more efficient application to the problem of object detection in images

For a particular recognition task of objects represented by a set of training images ($\boldsymbol{V}$) we need: (i) to calculate (in off-line mode) projection vectors of the training images onto the obtained NMF vector basis ($\boldsymbol{W}$) (feature vectors), and (ii) for each unknown input vector $\boldsymbol{y}$ to calculate (in on-line mode) a projection vector onto the obtained vector basis ($\boldsymbol{W}$). Guillamet & Vitrià (2003) proposed to use the feature vectors determined in the NMF run, i.e., the columns of matrix $\boldsymbol{H}$. The problem of determining projected vectors for new input vectors in a way that they are comparable with the feature vectors is solved by the authors by re-running the NMF algorithm. In this second run they keep the basis matrix $\boldsymbol{W}$ constant and the matrix $\boldsymbol{V}_{test}$ contains the new input vectors instead of the training image vectors. The results of the second run are the searched projected vectors in the matrix $\boldsymbol{H}_{test}$. However, this method has some weakness, that we described in Soukup & Bajla (2008) using an example of 3D point data instead of high-dimensional images. The points have been divided into two classes A, B, based on point proximity. We ran NMF to get a two dimensional subspace spanned by two vectors w_1 and w_2, which together build matrix $\boldsymbol{W}$. For each class, it can be observed that the projection rays are all non-orthogonal w.r.t. the plane and that their mutual angles significantly differ (even for feature vectors belonging to the same class). Thus the feature vectors of the set A and set B are not separated clusters anymore. We suspect that a reliable classification based on proximity of feature vectors could be achieved in this case (Soukup & Bajla, 2008).
A second possibility to determine proper feature vectors for an NMF subspace, which is conventionally used (e.g., mentioned in Buciu (2007)), is to re-compute entirely new training feature vectors for the classification phase by orthogonally projecting the training points (images) onto to NMF subspace. Unknown input data to be classified are similarly orthogonally projected to the subspace. It can be noticed that the feature vectors determined in this way preserve a separation of the feature vector clusters, corresponding to the cluster separation in the original data space. In view of these observations, we proposed to favor the orthogonal projection method (Soukup & Bajla, 2008).
Nonetheless, *both* methods have their disadvantages. The method of Guillamet and Vitrià operates with non-orthogonally projected feature vectors that directly stem from the NMF algorithm and do not reflect the data cluster separation in the subspace. On the other hand, the conventional method does not accommodate the optimal data approximation result determined in NMF, because one of the two optimal factor matrices is substituted by a different one in the classification phase. In Soukup & Bajla (2008), we proposed to combine the benefits of both methods into one, i.e., benefits of orthogonal projections of input data and

preservation of the optimal training data approximation of NMF. We achieve this by changing the NMF task itself. Before a brief presentation of this modification, we recall in more details how the orthogonal projections of the input data are computed.
As the basis matrix $\boldsymbol{W}$ is rectangular, matrix inversion is not defined. Therefore one has to use a pseudo-inverse of $\boldsymbol{W}$ to multiply it from the left onto $\boldsymbol{V}$ (compare with Buciu (2007)). Orthogonal projections of data points y onto a subspace defined by a basis vector matrix $\boldsymbol{W}$ are realized by solving the following overdetermined equation system:

$$\boldsymbol{W} \cdot \boldsymbol{b} = \boldsymbol{y} \tag{5}$$

for the coefficient vector $\boldsymbol{b}$. This can, for instance, be achieved via the Moore-Penrose (M-P) pseudo-inverse $\boldsymbol{W}^{\dagger}$ giving the result for the projection as

$$\boldsymbol{b} = \boldsymbol{W}^{\dagger} \cdot \boldsymbol{y}. \tag{6}$$

Similarly, for the NMF feature vectors (in the off-line mode) we determine $\boldsymbol{H}_{LS} = \boldsymbol{W}^{\dagger}\boldsymbol{V}$, where $\boldsymbol{H}_{LS}$ are projection coefficients obtained in the least squares (LS) manner. These coefficients can differ severely from the NMF feature vectors implicitly given by $\boldsymbol{H}$. It is important to state that the entries of $\boldsymbol{H}_{LS}$ can contain negative values.
If one has decided to use the orthogonal projections of input data onto the subspace as feature vectors, the fact that the matrix $\boldsymbol{H}$ is not used anymore in the classification phase and that the used $\boldsymbol{H}_{LS}$, that is a substitute for $\boldsymbol{H}$, is not nonnegative anymore, gives rise to the questions whether matrix $\boldsymbol{H}$ is necessary in NMF at all and whether the corresponding encoding coefficient necessarily has to be nonnegative. Moreover, using the orthogonal projection method, we do not make use of the optimal factorization achieved by NMF, as the coefficient matrix is altered for classification. Consequently, we proposed the following modification of the NMF task itself: *given the training matrix* $\boldsymbol{V}$, *we search for a matrix* $\boldsymbol{W}$ such that

$$\boldsymbol{V} \approx \boldsymbol{W} \cdot (\boldsymbol{W}^{\dagger} \cdot \boldsymbol{V}). \tag{7}$$

Within this novel concept (modified NMF), $\boldsymbol{W}$ is updated in the same way as in common NMF algorithms. Even the sparseness of $\boldsymbol{W}$ can be controlled by the standard mechanisms, e.g., those of Hoyer's method. Only the encoding matrix $\boldsymbol{H}$ is substituted by the matrix $\boldsymbol{W}^{\dagger} \cdot \boldsymbol{V}$ to determine the current approximation error. The modified NMF method with Hoyer's sparseness scheme (henceforth we will speak about the Modified Hoyer NMF method) is used in Section 5 for comparison to the proposed NMF method and to Lee-Seung NMF method.

4. A particular concept of the parts-based NMF subspace representation using subtemplates

4.1 Conceptual considerations

In the description of the recent results achieved in the area of NMF methods, provided in the introduction, the emphasis was put on problems occurring in applications of these subspace representation methods to image recognition tasks with occluded objects. We see these problems at two basic levels, (i) in methodological lack of the parts-based principle definition, and (ii) in insufficient systematic evaluation (in the relevant literature) of recognition of images with occlusions. In this section we will address the first point, whereas Section 5 is devoted to the second one.
In the papers dealing with applications of NMF to image recognition tasks, the concept of parts-based representation is considered on an intuitive level, some analogy to the results of neurology is only mentioned. Therefore we based our reasoning about applicability

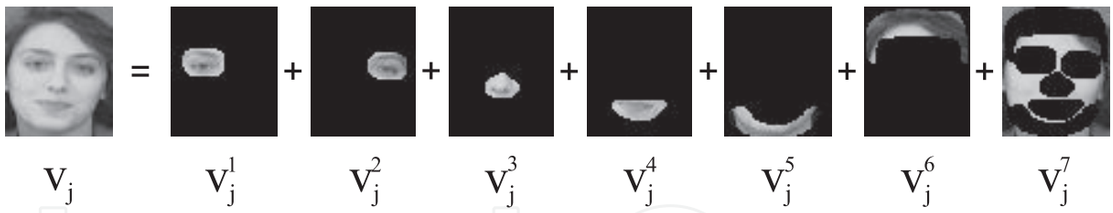


Fig. 1. Illustration of subtemplates for a face template with $p = 7$ parts defined.

of this principle to the NMF tasks on Hoyer's claims and Spratling's arguments. Both authors expressed a critical view to published statements that NMF subspaces follow a parts-based principle. They showed that single positive linear combination of NMF basis vectors obtained by conventional methods (considered as an analogy to combination of image parts) is not sufficient for achieving acceptable recognition rates in cases with occluded objects. They proposed some improvements; Hoyer by the introduction of a mechanism of explicit controlling of the matrix sparseness into the NMF scheme, Spratling by the development of an alternative dendritic inhibition neural network. The analysis of their results lead to the question: what property should NMF vector basis have, in order to really reflect the parts-based principle? We concluded that such a vector basis, in which the groups of vectors would uniquely correspond to individual image parts, could yield truly parts-based representation.

For his argumentation of discrepancy between the unsatisfactory results of application of the NMF to image recognition tasks with object occlusions and expectations of the parts-based representation bounded to the NMF methods, Spratling used a benchmark task of "bars-problem" with simple elementary image parts (bars). If the parts-based principle is tractable within the NMF representation of images with occlusions, then for any case of images with intuitively clear parts, the separate representation of parts by the corresponding NMF basis vectors should provide better approximation than the single nonnegativity of vector combinations. Such an NMF image representation should consequently lead to an improvement of occluded object recognition. Thus, our goal is to propose a benchmark task comprising real complex images which are composed of intuitively clear parts and to derive an NMF vector basis with basis vectors separately encoding these image parts.

4.2 Modular NMF

Intuitively clear understanding of *parts* of an image I_m can be based on the notion of set partition, namely, the partition of a set of raster points into a system of disjunctive subsets, the union of which is the whole raster. Similarly to the requirements used in image segmentation, we should consider only such subsets which give unique correspondence to individual objects or semantically unambiguous parts of an imaged reality (e.g., for face image we can consider as parts: left eye, right eye, nose, mouth, chin, and forehead with hair). In Fig. 1 an example is illustrated with the partition of a face into six parts and the face background as an extra part. For an image matrix I_m representing this face image (template) and subsets $P_1, P_2, \ldots, P_p$ which represent its individual parts we can write $I_m = P_1 + P_2 + \ldots + P_p$. Our intention is to formulate separate NMF tasks for the given parts of an image I_m that, however, would have data structure consistent with the initial NMF task. This means to preserve matrix size (n, m) of the initial image matrix I_m. We propose to do it by definition of matrices with the size (n, m) identical to the size of I_m in which all entries, except those corresponding to the given subset P_j, $j = 1, 2, \ldots, p$, are set to zeros. In accordance with the notation used in the domain of NMF methods, for the input matrix $\boldsymbol{V}$ with the columns $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$, which represent

template images, we denote individual subtemplates of the j-th part as $\mathbf{v}_1^j, \mathbf{v}_2^j, \ldots, \mathbf{v}_m^j$. Thanks to the identical sizes of subtemplate image matrices we can express m templates from the input matrix $\boldsymbol{V}$ as the sums of the corresponding subtemplates:

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{v}_1^1 + \mathbf{v}_1^2 + \ldots + \mathbf{v}_1^p, \\ \mathbf{v}_2 &= \mathbf{v}_2^1 + \mathbf{v}_2^2 + \ldots + \mathbf{v}_2^p, \\ &\vdots \\ \mathbf{v}_m &= \mathbf{v}_m^1 + \mathbf{v}_m^2 + \ldots + \mathbf{v}_m^p \, . \end{aligned} \tag{8}$$

Instead of one NMF problem for the input template matrix $\boldsymbol{V}$ of type $(n \times m)$, the basis vector matrix $\boldsymbol{W}$ of type $(n \times r)$, and the matrix $\boldsymbol{H}$ of encoding coefficients of type $(r \times m)$,

$$\boldsymbol{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m] \approx \boldsymbol{W} \cdot \boldsymbol{H} \, , \tag{9}$$

we formulate p separate NMF tasks, each for the separate j-th part (i.e., for all subtemplates representing this part):

$$\boldsymbol{V}^j = [\mathbf{v}_1^j, \mathbf{v}_2^j, \ldots, \mathbf{v}_m^j] \approx \boldsymbol{W}^j \cdot \boldsymbol{H}^j \, . \tag{10}$$

The dimensions $r^\star$ of subspaces generated by each of p separate NMF tasks are identical and we define them as the integer part $r^\star = [r/p]$. For the j-th separate NMF task, $j = 1, 2, \ldots, p$, we express the $(n \times r^\star)$- matrix $\boldsymbol{W}^j$ of $r^\star$ basis column vectors as

$$\boldsymbol{W}^j = [\mathbf{w}_1^j, \mathbf{w}_2^j, \ldots, \mathbf{w}_{r^\star}^j] \, , \tag{11}$$

and, similarly, the $(r^\star \times m)$-matrix $\boldsymbol{H}^j$ of encoding coefficients -column vectors as

$$\boldsymbol{H}^j = [\mathbf{h}_1^j, \mathbf{h}_2^j, \ldots, \mathbf{h}_m^j] \, . \tag{12}$$

The individual components of the z-th column vector of this matrix are denoted as follows:

$$[\mathbf{h}_z^j]' = [h_{z1}^j, h_{z2}^j, \ldots, h_{zr^\star}^j] \, . \tag{13}$$

Assume we have solved all separate NMF tasks for p sets of subtemplates of image parts. Thus we have obtained p NMF subspaces described by the matrices $\boldsymbol{W}^1, \boldsymbol{W}^2, \ldots, \boldsymbol{W}^p$ of basis column vectors. For the j-th separate NMF task we can express the $z-$th column vector of the input data matrix $\boldsymbol{V}^j$ as an NMF-approximated linear combination of basis vectors:

$$\mathbf{v}_z^j \approx h_{z1}^j \mathbf{w}_1^j + h_{z2}^j \mathbf{w}_2^j + \ldots + h_{zr^\star}^j \mathbf{w}_{r^\star}^j \, . \tag{14}$$

In matrix notation we get:

$$\mathbf{v}_z^j \approx [\mathbf{w}_1^j, \mathbf{w}_2^j, \ldots, \mathbf{w}_{r^\star}^j] \cdot \mathbf{h}_z^j \, . \tag{15}$$

We would like to express the z-th template, i.e., z-th column vector $\mathbf{v}_z$ of the input matrix $\boldsymbol{V}$ of the initial NMF task using the results of the separate NMF tasks. First, according to our partition scheme (8), we can express the column vector $\mathbf{v}_z$ as

$$\mathbf{v}_z = \mathbf{v}_z^1 + \mathbf{v}_z^2 + \ldots + \mathbf{v}_z^p \, . \tag{16}$$

Using the results of the solution of p separate NMF tasks (using e.g., the Lee and Seung optimization scheme for the L_2-norm as a cost function) given in (15), we obtain the following (NMF) approximation of the given template

$$\begin{aligned}\mathbf{v}_z \approx [\mathbf{w}_1^1, \mathbf{w}_2^1, \ldots, \mathbf{w}_{r^\star}^1] \cdot \mathbf{h}_z^1 + [\mathbf{w}_1^2, \mathbf{w}_2^2, \ldots, \mathbf{w}_{r^\star}^2] \cdot \mathbf{h}_z^2 + \\ \ldots + [\mathbf{w}_1^p, \mathbf{w}_2^p, \ldots, \mathbf{w}_{r^\star}^p] \cdot \mathbf{h}_z^p \ .\end{aligned} \tag{17}$$

The latter formula can be re-written in matrix notation

$$\mathbf{v}_z \approx [\mathbf{w}_1^1 \ldots \mathbf{w}_{r^\star}^1, \mathbf{w}_1^2 \ldots \mathbf{w}_{r^\star}^2, \ldots, \mathbf{w}_1^p \ldots \mathbf{w}_{r^\star}^p] \cdot \begin{bmatrix} \mathbf{h}_z^1 \\ \mathbf{h}_z^2 \\ \vdots \\ \mathbf{h}_z^p \end{bmatrix},$$

where components of the subcolumns in the matrix $\boldsymbol{H}$ are given in (13). For the whole input matrix $\boldsymbol{V}$ of templates we obtain as an approximative equality, the result of an optimization task of the NMF problem:

$$\begin{aligned}\boldsymbol{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m] \approx \\ \approx [\boldsymbol{W}^1, \boldsymbol{W}^2, \ldots, \boldsymbol{W}^p] \cdot \begin{bmatrix} \mathbf{h}_1^1, \mathbf{h}_2^1, \ldots, \mathbf{h}_m^1 \\ \mathbf{h}_1^2, \mathbf{h}_2^2, \ldots, \mathbf{h}_m^2 \\ \ldots \\ \mathbf{h}_1^p, \mathbf{h}_2^p, \ldots, \mathbf{h}_m^p \end{bmatrix} = \\ = [\boldsymbol{W}^1, \boldsymbol{W}^2, \ldots, \boldsymbol{W}^p] \cdot \begin{bmatrix} \boldsymbol{H}^1 \\ \boldsymbol{H}^2 \\ \vdots \\ \boldsymbol{H}^p \end{bmatrix} = \boldsymbol{W}^\star \cdot \boldsymbol{H}^\star \ .\end{aligned} \tag{18}$$

The approximate factorization of the input image template matrix $\boldsymbol{V}$ obtained in this way comprises basis vectors which uniquely correspond to the individual image parts defined in terms of a set of subtemplates of these parts. A remaining question is: what is the relation between our separated factorization $\boldsymbol{W}^\star \cdot \boldsymbol{H}^\star$, based on the Modular NMF, and any original factorization $\boldsymbol{W} \cdot \boldsymbol{H}$?
Let us denote $reserr_k$ the residual L_2-error of the k-th separate NMF task:

$$reserr_k = \sum_{i,j} \left\| V_{i,j}^k - (\boldsymbol{W}^k \boldsymbol{H}^k)_{ij} \right\|^2 \ .$$

Since in each individual NMF problem we solve the separate optimization problem (in L_2 norm) which differs from that formulated for the entire image, the residual error of the separated factorization of the entire image is not equal to the sum of the residual errors for the

individual NMF tasks:

$$\|\boldsymbol{W}^{\star}\boldsymbol{H}^{\star} - \boldsymbol{V}\|^2 \neq \sum_k reserr_k = \\ = \sum_k \sum_{i,j} \left\| V^k{}_{i,j} - (\boldsymbol{W}^k \boldsymbol{H}^k)_{ij} \right\|^2 .$$

Neither the equality between the residual error of an NMF solution of the original entire image $||\boldsymbol{WH} - \boldsymbol{V}||^2$ and the residual error of our separated factorization $||\boldsymbol{W}^{\star}\boldsymbol{H}^{\star} - \boldsymbol{V}||^2$ is valid. What we can do is to formulate a modified optimization NMF problem in L_2 norm, as it is given in Eq. (2), with the initial matrices $\boldsymbol{W}^{\star}$ and $\boldsymbol{H}^{\star}$ instead of the matrices $\boldsymbol{W}, \boldsymbol{H}$, initialized by random entries. According to Lee & Seung (1999), the convergence property is maintained with all initial values of $\boldsymbol{W}$ and $\boldsymbol{H}$, only resulting optimum may be altered. Thus, in our case, we use $\boldsymbol{W}^{\star}$ and $\boldsymbol{H}^{\star}$ to drag the NMF algorithm into the desired direction of the parts-based representation.

5. Computer experiments – a comparative study

5.1 Goals

Our analysis and exploration in the previous sections can be summarized in the following way. We have documented that the application of the conventional NMF method of Lee and Seung to image recognition problems with object occlusions does not provide expected parts-based representations. The further attempts to improve applicability of NMF to the recognition of occluded image objects, resulted in various NMF modifications. The semi-NMF approach, we proposed in Soukup & Bajla (2008) as a modification of Hoyer's NMF algorithm, manifested higher recognition rates for some occluded cases of the ORL face database. However, due to the acceptance of negative terms in the linear combination of the obtained NMF basis vectors, the method is even more distant to the parts-based principle. Based on this finding, our next intention was to modify the NMF scheme towards a vector subspace representation that is more compatible with the parts-based principle. The novel Modular NMF algorithm, we have proposed in the previous section, represents a possible improvement in this direction. The basic goal of the computer experiments was to explore behavior of these three NMF algorithms under various conditions. A detailed comparative study should contribute to the explanation of several unclear aspects we encountered in the papers on NMF in which the suitability of the NMF for image object recognition with occlusions was advocated. Moreover, as during some preliminary tests, reported in this Section, it appeared that using the conventionally used face databases suffers from some methodological drawbacks, we decided to analyze first the correctness of the test images and thereby to ensure the unified reference basis for comparisons. The details will be given below.

5.2 Testing conditions and our revisions of input data

There are five key aspects (variables, parameters) which can affect the recognition rate of the NMF algorithms applied to the given problem:

1. type and resolution of the images used for recognition,
2. type of partial object occlusions and the method of their detection and suppression,
3. classification method used,
4. metric of the NMF vector subspace used,
5. dimension of the NMF vector subspace chosen.

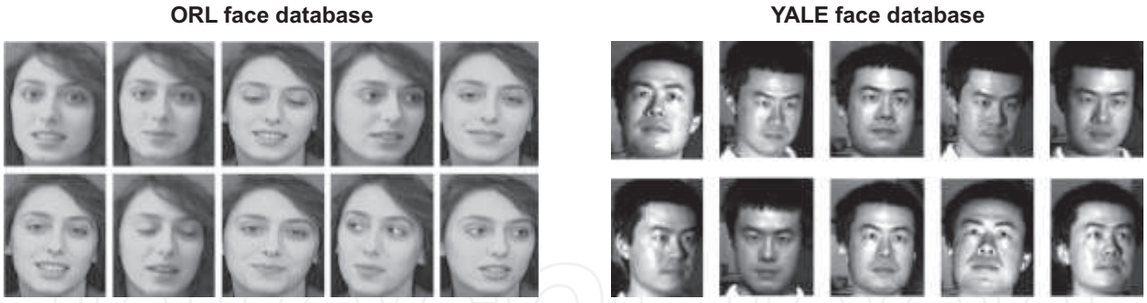


Fig. 2. Examples of face images selected from the ORL and YALE face databases.

5.2.1 Image databases

In our computer experiments we needed appropriate image databases with images containing intuitively clear parts. The public databases with faces satisfied this requirement. As in our computational study (Bajla & Soukup, 2007) we used 222 training images, and 148 testing images, selected from 370 faces of the Cambridge ORL face database[1] (Fig. 2, left). These two sets of images were chosen as disjunctive sets in a standard ratio of 60% for the training set and 40% for the testing set. The gray-level images with resolution 92×112 have been downsampled to the resolution $46 \times 58 = 2668$ pixels. For these face images, we defined four intuitively apparent parts of the face: left eye, right eye, nose, and mouth. The fifth part was determined as a complement of the union of all four face parts.

The second gray-level image database of faces we have selected is the YALE B face database[2] (Fig. 2, right). The database contains 5760 single light source images of 10 subjects, each seen under 576 viewing conditions (9 poses and 64 illumination conditions). For our experiments we have limited the number of illumination conditions to 5 representative cases, so that each subject was represented by 45 images. For each person, we have selected 31 images as training and 14 as testing, getting altogether the training set with 310 face images and the testing set with 140 face images. The resolution of the images is $62 \times 82 = 5084$ pixels.

In our preceding computer experiments (the study Bajla & Soukup (2007) and the paper Soukup & Bajla (2008)), also the CBCL face image database, comprising gray-level images with resolution of $19 \times 19 = 361$ pixels has been used[3]. Since the resolution of this database is much lower than in case of images from the two previously mentioned databases, for preserving approximately equivalent conditions, we have decided not to consider this database in the experiments reported in this study.

In the field of image object recognition, in particular, in the tasks in which the face image databases are standardly used, recently in Ling et al. (2006); Shamir (2008) a suspicion appeared that various classifiers, explored in these tasks, exploit not only the relevant information (face pixels), but they considerably utilize also an additional information contained in object background, implicitly comprised in most of the face images. If so, it should have significant consequences on correctness of a unified reference basis of testing image data for evaluation of performance of classifiers in image object recognition. Therefore, it was necessary to examine this suspicion on the selected two face image databases.

To address this question, we introduced alternative training and testing data sets (called "cropped") that contain no background pixels. The background was eliminated by cropping each image tightly around all the known face parts (i.e., left eye, right eye, nose, mouth).

[1] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[2] http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html
[3] http://cbcl.mit.edu/cbcl/software-datasets/FaceData2.html

Later on, the cropped images were resampled to their original size by a bilinear interpolation. A result of such a cropping operation is shown in Fig. 3 (middle row). Note that, besides the apparent background elimination, this operation partly also normalizes positions of individual face parts.
Related to our analysis of the role of the parts-based principle used in applications of the NMF approaches to image object recognition, we made in Section 2, another aspect had to be investigated, namely, a possibility of how to transform input data in order to normalize geometrical location of the face parts in training and testing images. It can be shown that the NMF method may arrive at the parts-based representations only when distinguished image parts (i.e., principal building blocks) either reside at approximately stable positions or their shape does not vary too much. Otherwise the NMF optimization algorithm is not capable of finding a low dimensional subspace basis that would capture both shape and position variations of possibly occurring image patterns. In case of the face images from the ORL and YALE databases, one can observe significant differences between shapes of multiple parts of the same kind (e.g., eyes of different individuals), as well as variability in their placement within an image (Fig. 3, top row). This is an additional evidence supporting the idea that the raw facial data without any initial adjustment are not suitable for the NMF processing.
To answer this question, we proposed yet another training and testing data sets (called "registered") which contain faces with normalized positions of the face parts. Every image was transformed by an affine transformation so that centroids of its parts approximate the predefined positions as much as possible (Fig. 3, bottom row). Note that the centroid distribution here spreads much less than in the cases of the original and cropped data.

5.2.2 Occlusions

The topic of modeling image object occlusions has not been yet systematically addressed in the NMF literature. In computer experiments with various NMF methods applied to images from three image databases (Bajla & Soukup, 2007), we observed that recognition rates are, in general, sensitive to the location of occlusions. In order to examine the parts-based principle within the NMF scheme, the following aspects of object occlusions are of our interest:

- occlusions should have unique relations to natural facial parts,
- images containing artificial occlusions should be still recognizable by a human observer.

Our preliminary experiments showed that the exact geometrical shape of an occlusion does not influence the obtained RR values as much as its position. Consequently, we have decided for two alternatives, simple rectangles and more detailed polygonal regions covering the individual facial parts. The regions were defined manually by hand for both training and testing data within both ORL and YALE databases, however, only in the case of the Modular NMF, this information was used within the training process. To characterize behavior of projections of occluded faces in NMF subspaces and to evaluate the recognition results on a systematic basis, we have generated a system of four elementary facial occlusions: left eye, right eye, nose, and mouth (Fig. 4, 1-4). Additionally, four complex occlusion types have been defined as combinations of some of the four elementary occlusions (Fig. 4, 5-8).
To simulate a severe facial occlusion we have decided to replace original pixel intensities by zero values. As the assumed face parts have typically a higher brightness, their unoccluded pixel intensities range normally in higher values. Thus, in L_2 sense, the zero values within the occlusions tend to shift the occluded images far from their original unoccluded version (here the images are represented as vectors). Furthermore, such occlusions simulate presence of typical real world occlusions such as mustache or sunglasses, etc (Fig. 4, left). Hereinafter we call this occlusion type as "black occlusion".

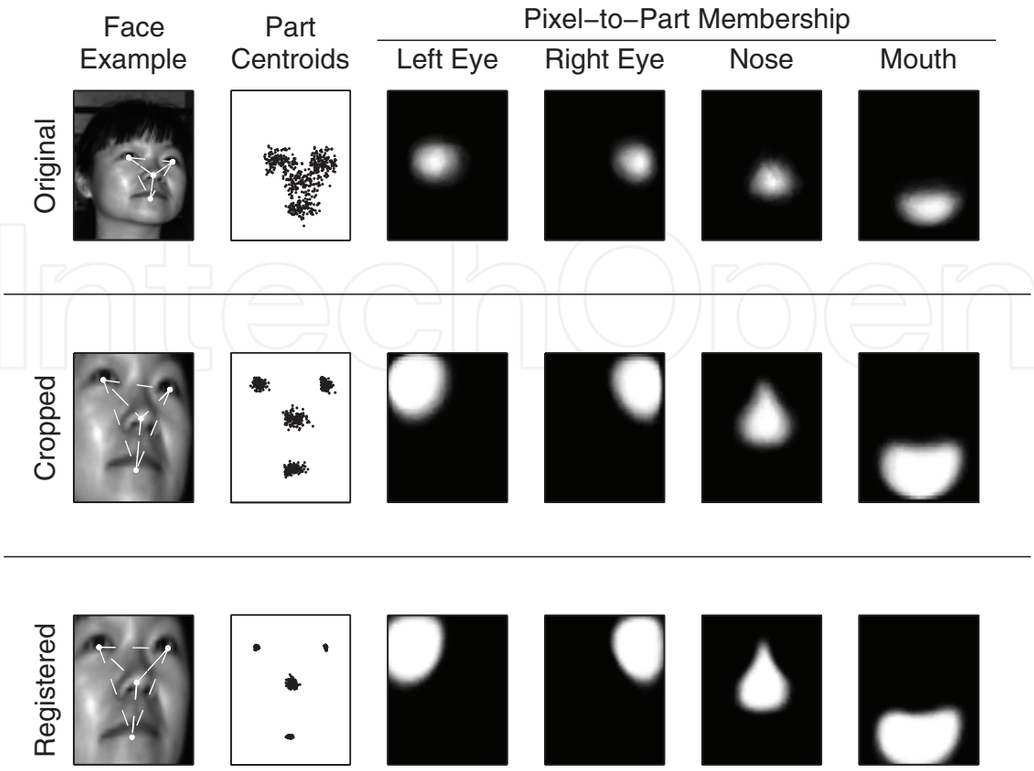


Fig. 3. An example of the original, cropped and registered face from the YALE database. The plots in the leftmost column contains original and transformed face images with centroids of individual face parts (i.e., left eye, right eye, nose, mouth) marked by white points. The second column comprises plots of distributions of the part centroid positions. The remaining four columns show plots of the pixel-to-part membership functions for different face parts. These functions express an estimated probability that particular pixel belongs to the certain face part.

If there exists a method for detecting the area associated with an occlusion, one may pose a question how to incorporate this a priori information into the NMF algorithms. In general there are two possible approaches to this problem: i) to suppress intensities belonging to the occlusion and replace them with values from the normal facial range, or ii) completely exclude the occluded image pixels from NMF calculations.

As for the first approach, we have implemented the occlusion suppression idea by filling the occluded image pixels by values interpolated from the nearest unoccluded pixels (Fig. 4, right). Such corrected images were treated the same way as occluded images and no further modifications of the NMF algorithms were required. Hereinafter we call this occlusion type as "interpolated occlusion".

As mentioned above, the second method for suppressing known occlusions is based on elimination of the occluded image pixels from the entire NMF calculations. Since the occlusion can only occur in the classification (testing) phase (i.e., the training data cannot be disturbed by any occlusions), the NMF training algorithms remain the same, however, a slight modification of the NMF classification procedure is required. Assuming that the exact position and extent of the occlusion is a priori known for every classified image, one can mask out (replace by zeros) all the occluded pixels in the classified image, as well as in all the training images. Only then these images are projected onto the NMF subspace and the classification algorithm is

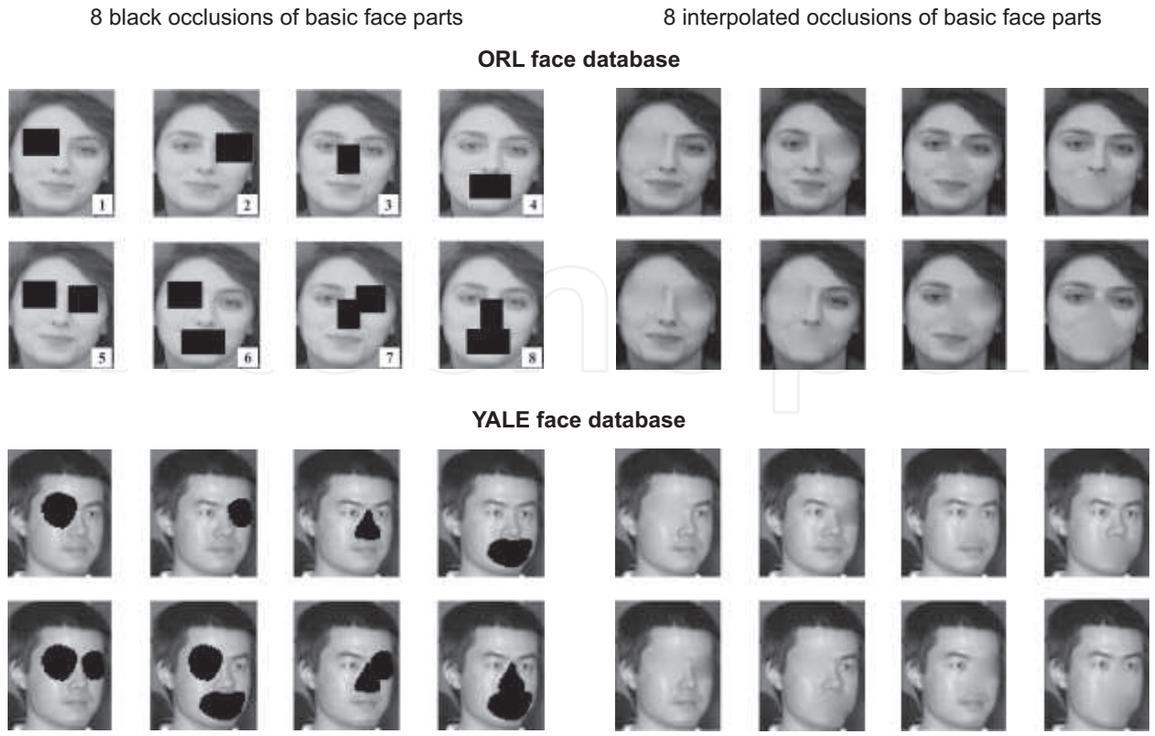


Fig. 4. Test face images with various types of partial occlusions - the left column - full black occlusions, and the right column - interpolated occlusions. The first row of images in each database example represents the elementary occlusions, while in the second row combined occlusions are displayed.

applied to the obtained feature vectors. Hereinafter, we call this occlusion type as "masked occlusion".

Understanding the essence of the mentioned three occlusion types (i.e., black, interpolated, and masked) and their connection to the NMF principles, one can make several assumptions about the performance of NMF applied to such data. Since the black occlusions represent the case where invalid pixels belonging to an occlusion have most severe impact on the final classification decision, the lowest performance should be expected here. In the case of the interpolated occlusions, where the impact of the occluded pixels is already partially suppressed, the classification performance should be significantly better than with the black occlusions. The best performance should be expected in the case of the masked occlusions where all disturbing image pixels are correctly excluded from the calculation and only the remaining valid information is used for making classification decision. Nevertheless, the need of having available all the original training data during the testing phase and the necessity of their masking and projecting along with every new classified image makes this approach highly memory and computationally demanding and, thus, rather difficult to apply in real situations.

5.2.3 Metrics

For the vector subspace methods of image object recognition a metric should be specified first. It measures the distance between the projection of the image being processed onto the subspace and the projections of the training images onto the same subspace (these are usually called feature vectors). In Bajla & Soukup (2007) we reported the results of vast

experiments with various types of metrics: Euclidean, Riemannian (Guillamet & Vitrià, 2003), "diffusion-like" (Ling & Okada, 2006), and our modified Riemannian metric, carried out for various image databases and for several NMF methods. These results showed that Recognition Rates (RR) for Euclidean metric are comparable to those obtained for Riemannian metric and that they are much higher than expected in the paper of Guillamet & Vitrià (2003), and Liu & Zheng (2004). Therefore in our experiments with occluded faces we have used only Euclidean metric.

5.2.4 Classifiers

For recognition experiments with subspaces generated by the individual NMF approaches, it is necessary to classify each projection of a test image onto a NMF subspace in classes represented by feature vectors of the given subspace. The classification can be accomplished by means of various approaches, but usually the Nearest Neighbor Classifier (NNC) is used. The NNC is a standard stable non-parametric classifier providing good results having sufficient number of training examples. Some of the good properties of NNC are summarized in Duda et al. (2001).

In preliminary experiments with both face databases we tested also the classifier that utilizes the information on mean centers of the classes of feature vectors (MCC). The results obtained showed that for all three NMF methods, and for both basic types of occlusions (black or interpolated), as well as for all individual cases of partial occlusions (elementary and combined), the RR values achieved for the MCC have been significantly lower than the RR values for the NNC.

Testing more sophisticated classifiers in our experiments of NMF was beyond our interest, but more importantly, the main reason of the application of the NNC exclusively, was to ensure the unified methodological basis of comparison used in most of the papers published in this area (see References).

Subspace dimensions

For both face image databases, we have varied the dimensions of the NMF-subspaces from $r = 25$ up to 250. Elementary and combined types of occlusions have been applied to all test images. The recognition rates have been calculated separately for each set of testing faces with one occlusion type.

For each of the training face images we determined regions of the defined parts by hand (using approximate curvilinear boundaries) which served as subtemplates of the given database used in the Modular NFM algorithm. The sets of training subtemplates have been used for learning the basis image vectors of the individual separate NMF tasks which have been accomplished using Hoyer's algorithm with the controlled sparseness s_W. For solving the Modular NMF problems for the entire image we have used again the algorithm of Hoyer.

In Fig. 5 examples of basis vectors of 140-dimensional NMF subspaces are illustrated. These have been generated by solving two conventional NMF methods for the ORL face image database: Lee-Seung NMF method (Fig. 5, left), and the localized Modular NMF method (Fig. 5, right). Whereas the global nature of the conventional NMF basis vectors is apparent from Fig. 5, left, the vector bases of the Modular NMF method with separated vectors of individual face parts, clearly manifest locality of basis vectors (Fig. 5, right).

5.3 Benchmarking of three NMF methods applied to two face databases and using two types of image partial occlusions

The inclusion of the three above mentioned NMF variants pursued the basic goal to demonstrate their different performances in recognition tasks with occluded image objects.

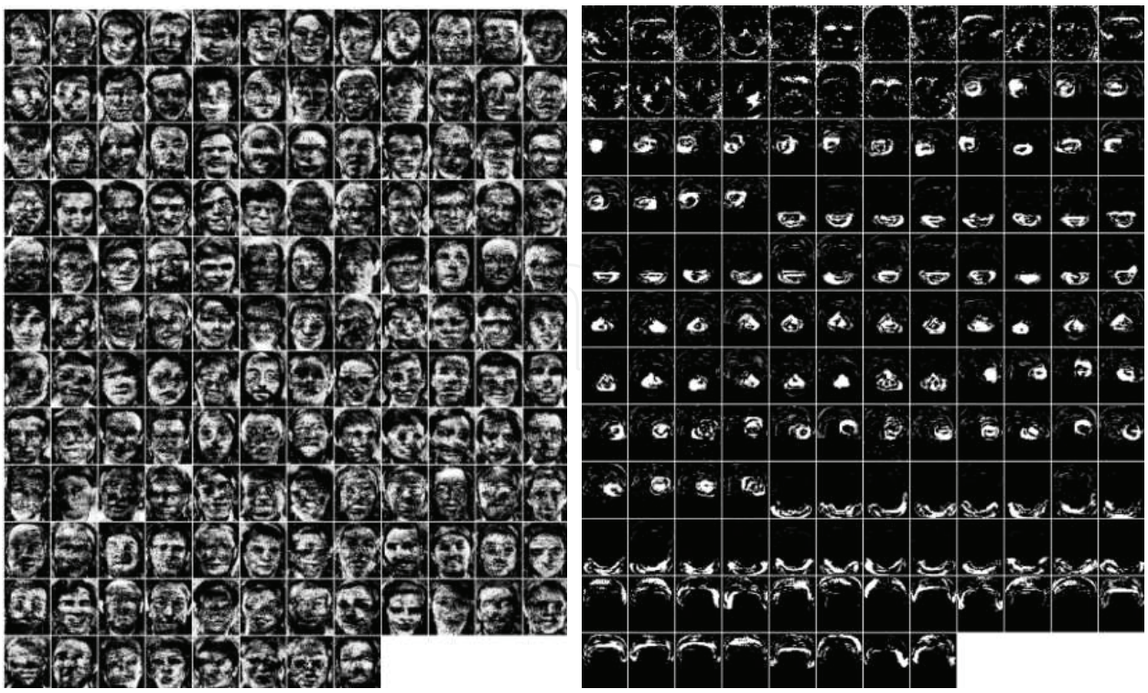


Fig. 5. Visualization of the vector bases of the 140-dimensional image subspaces for the ORL database: i) for the conventional Lee-Seung NMF method (left), ii) for the localized Modular NMF method (right).

The algorithmic versions differ in the following methodological characteristics: (i) the Lee-Seung NMF algorithm provides strictly nonnegative matrix factorization, however, its application to images reveals a discrepancy between the motivating parts-based principle and the real shape of subspace vector basis, and moreover, its expected overall superiority in recognition rates was not confirmed in practice, (ii) the Modified Hoyer NMF algorithm is not strict nonnegative (it is a semi-NMF method), it manifests no apparent relation to the parts-based principle, (iii) the proposed Modular NMF algorithm provides strictly nonnegative matrix factorization and it very closely reflects the parts-based principle. Thus each individual experiment includes the RR values (ordinate) for these three NMF versions which are graphically discriminated in the plots. As the basic variable parameter (abscissa) of the recognition, the NMF subspace dimension is used (Fig. 6).

5.3.1 Evaluation

As a basis for comparison, the NNC of the face images has been performed in the original data space without application of the NMF methods. In Table 1 we document the obtained RR values which confirm the claim of some researchers that questioned the suitability of using raw images contained in standard public domain face image databases. As mentioned above, we have examined this problem on two databases, i.e., ORL and YALE. First of all, a decrease of the RR values should be noted for cropped and registered unoccluded data (numbers listed in the parentheses in the second row of the table) in comparison to RR obtained for the nontransformed original input data. Further, the results obtained for occluded images show that for the raw input data, only minor RR differences between black and interpolated occlusion types are observed (about 9%). In contrast to these characteristics, the RR values achieved for the cropped and registered data manifest a radical change (about 42%). The most

	Original data				Cropped data				Registered data			
	ORL (.958)		**YALE** (.956)		**ORL** (.791)		**YALE** (.713)		**ORL** (.813)		**YALE** (.802)	
Occlusion	*Black*	*Interp.*	*Black*	*Interp.*	*Black*	*Interp.*	*Black*	*Interp.*	*Black*	*Interp.*	*Black*	*Interp.*
Left Eye	.895	.965	.889	.926	.194	.736	.279	.588	.201	.715	.308	.669
Right Eye	.909	.951	.926	.933	.319	.750	.500	.691	.298	.763	.522	.750
Mouth	.854	.951	.882	.948	.138	.770	.323	.705	.145	.770	.264	.742
Nose	.923	.951	.941	.955	.361	.763	.698	.757	.416	.777	.713	.823
LE+RE	.826	.958	.772	.911	.111	.631	.139	.441	.138	.534	.176	.485
LE+Mo	.798	.958	.772	.926	.118	.687	.117	.558	.097	.694	.114	.588
No+Mo	.756	.951	.852	.955	.076	.687	.338	.705	.083	.666	.294	.742
RE+No	.854	.944	.867	.955	.152	.729	.367	.705	.131	.701	.470	.727

Table 1. The RR values obtained by NNC operating in the original image space (i.e., no NMF applied). Different occlusion types, as well as no occlusions (numbers in parentheses) have been considered.

significant decrease of RR values (up to 10 times in the case of the occlusion combination of the nose and mouth) can be observed for cropped ORL images with black occlusions. RR decrease for the same types of occlusions in the case of YALE images is not so extreme. RR obtained for the interpolated types of occlusions are for both databases with cropped face images significantly higher than for the black occlusions. Some improvement in RR increasing is reached for the registered input images in all tested cases.

A set of plots of RR values versus NMF subspace dimensions are depicted in Fig. 6. In these figures we outline a complex picture of the research results on image data representation and reduction using the NMF methods. The ensemble of plots reflects the individual aspects of how the parts-based principle is reflected in each particular combination of parameters. We addressed these aspects in the previous sections. Using the mean RR values, calculated over all elementary types of occlusions (represented in the plots by circles), we intend to express tendencies prevailing in the tested face image recognition. Moreover, this information is completed by gray stripes depicting intervals between minimum and maximum RR values obtained for the elementary occlusions. Fig. 6a shows the results obtained for the ORL face image database, whereas Fig. 6b shows the results for the YALE database. For each database, square blocks of nine plots are related to different NMF approaches (i.e., Lee-Seung NMF, Modified Hoyer NMF, Modular NMF). Nine plots ordered in a single row summarize the RR values achieved for one type of input data (as explained above, we used original, cropped, and registered data). As for three basic types of occlusions (i.e., black, interpolated, and masked) included in our experiments, the plots belonging to one of these types are always collected into a single column. Based on the detailed analysis of relations between the RR plots given in Figs 6, the following findings can be formulated:

- When using the original data, comprising the additional information on the face background, small differences in the mean RR values obtained for all NMF methods and occlusion types, are observed. These values are very close to the RR values obtained for unoccluded image objects (marked as thick gray curves). Apparent differences can be observed only for the representative case of the combined occlusions (marked as thin gray curves). This finding relates equally to ORL and YALE databases.
- As for the cropped face images from both databases, which represent a correct reference basis for benchmarking the individual NMF methods applied to various situations in our research, significant differences are observed. The decrease of RR can be observed already for unoccluded data.

Fig. 6. Integrated presentation of RR for all aspects of NMF methods explored in the course of this study. Two basic blocks of plots are displayed for the face images from ORL (a) and YALE (b) databases. Each individual plot represents the RR values (ordinate) for one NMF method, one basic type of occlusion and ten dimensions of NMF subspaces used in the experiments (abscissa). In the plots, the circles represent the mean RR values calculated over all elementary occlusions, while the gray stripes depict intervals between relevant minimum and maximum recognition rates. The thick gray curves stand for RR values obtained for unoccluded images, while the thin gray curves show recognition rates for representative case of the combined occlusions (LE+RE). The detached bar on the right side of each plot (cross-circle-cross) represent RR values for NNC operating in the original image space.

- The goal of the registration transformation of face image data was to preserve an approximate constant position of face parts over the sets of training, as well as testing images and thereby to provide data suitable for building a parts-based representation. To preserve the correct basis for benchmarking, we applied the registration transformation to the cropped input face images, then, the contribution of this transformation to the improvement of RR can be characterized by a slight increase of RR for both databases, both basic occlusion types and for all three NMF methods explored.
- Focusing our attention only to the registered input data and to the results related to black occlusions, which reflect a real recognition of the raw face images occluded by objects comprising strongly disturbing intensities (approximately zeros), the ability to recognize the face image differs for the individual NMF methods explored. Namely, in the case of the ORL data, the lowest RR mean values are reached for the Modified Hoyer method, and the highest RR mean values are obtained for the Modular NMF method, moreover, this method is characterized by the narrowest stripe of RR variance over elementary occlusion types. RR obtained for the worst combined occlusion case (LE+RE) are also maximal for the Modular NMF. In the case of YALE data, the lowest RR mean values, with apparently widest variance stripes, are achieved again by the Modified Hoyer method. The slightly higher RR mean values than for the Modular NMF method are obtained for the conventional Lee-Seung NMF method, note that considerably higher RR values than in the case of ORL data have been obtained.
- In the plots, it can be seen that considering the interpolated occlusions, representing real situations when intensities belonging to face parts are closer to their normal range (unoccluded), leads to significant improvement of RR for all NMF methods explored and for both image databases (here, we still limit our focus on the registered input data).
- It is worth mentioning that for the registered data, and the maximum data reduction (i.e., subspace dimensions 25-50), the highest RR values are reached for the Modular NMF method – independently on black or interpolated occlusions,
- As for the masked occlusions, the plots confirm our expectation about a maximum improvement of RR for all three NMF methods. In this case almost no difference can be seen between occluded and unoccluded face images.

The explicit particular RR values obtained for all recognition situations included in the computer experiments, from which plots in Fig. 6 have been constructed, are given in Tables 2 and 3. Due to space limitations we made a selection of the most representative RR values (subspace dimensions 25, 250, and Original space).

6. Conclusions

In this research study, using the relevant papers published in the given area, we have analyzed the relation of the parts-based principle to the methodology of NMF data representation when applied to computer vision tasks of image object recognition under partial object occlusions. Beginning by the conventional Lee-Seung NMF method (Lee & Seung, 1999), that does not comprise any explicit concept of a vector sparsity, neither yields subspaces with the vector bases corresponding to natural image parts, the NMF algorithm development in the NMF literature proceeded towards the explicit incorporation of the vector sparsity constraints into the NMF optimization problem, and towards more locally specified vector bases of the NMF representation. After the short description of the NMF algorithm, that we developed in Soukup & Bajla (2008), as a more numerically efficient modification of Hoyer's NMF algorithm (Hoyer, 2004), we have proposed the novel Modular NMF approach that preserves

Subspace dimension	No occ.	Black occ.				Interpolated occ.				Masked occ.			
		Max	Mean	Min	LE+RE	Max	Mean	Min	LE+RE	Max	Mean	Min	LE+RE
Lee-Seung NMF – Original (top) / Cropped (middle) / Registered (bottom) images													
25	.932	.891	.862	.816	.762	.932	.930	.925	.925	.939	.934	.925	.952
250	.871	.843	.821	.802	.715	.884	.872	.864	.884	.904	.887	.864	.898
Orig. space	.958	.923	.895	.854	.826	.965	.955	.951	.958	–	–	–	–
25	.655	.162	.100	.027	.068	.618	.596	.554	.463	.666	.638	.619	.622
250	.790	.385	.286	.144	.191	.802	.742	.694	.627	.788	.772	.755	.703
Orig. space	.791	.361	.253	.138	.111	.770	.755	.736	.631	–	–	–	–
25	.723	.210	.185	.162	.096	.700	.645	.581	.511	.754	.709	.668	.625
250	.805	.418	.346	.279	.150	.775	.738	.682	.559	.823	.801	.763	.688
Orig. space	.813	.416	.265	.145	.138	.777	.756	.715	.534	–	–	–	–
Modified Hoyer NMF – Original (top) / Cropped (middle) / Registered (bottom) images													
25	.932	.883	.806	.735	.627	.959	.939	.898	.898	.965	.950	.939	.966
250	.945	.931	.899	.850	.837	.952	.948	.945	.959	.952	.952	.952	.952
Orig. space	.958	.923	.895	.854	.826	.965	.955	.951	.958	–	–	–	–
25	.668	.162	.101	.027	.068	.625	.602	.540	.470	.659	.639	.628	.642
250	.783	.400	.274	.202	.097	.802	.740	.714	.566	.802	.767	.736	.737
Orig. space	.791	.361	.253	.138	.111	.770	.755	.736	.631	–	–	–	–
25	.642	.216	.125	.074	.048	.639	.582	.533	.422	.652	.629	.614	.604
250	.777	.406	.239	.121	.077	.761	.699	.614	.518	.822	.767	.729	.688
Orig. space	.813	.416	.265	.145	.138	.777	.756	.715	.534	–	–	–	–
Modular NMF – Original (top) / Cropped (middle) / Registered (bottom) images													
25	.837	.808	.755	.673	.572	.863	.852	.843	.836	.884	.852	.836	.858
250	.939	.932	.928	.925	.925	.945	.94	.938	.945	.952	.943	.938	.939
Orig. space	.958	.923	.895	.854	.826	.965	.955	.951	.958	–	–	–	–
25	.783	.509	.381	.319	.118	.741	.696	.632	.491	.815	.764	.743	.723
250	.709	.530	.502	.476	.239	.686	.663	.632	.551	.700	.684	.655	.662
Orig. space	.791	.361	.253	.138	.111	.77	.755	.736	.631	–	–	–	–
25	.758	.448	.361	.291	.171	.781	.696	.648	.551	.768	.743	.723	.653
250	.738	.557	.522	.479	.314	.755	.706	.659	.518	.768	.752	.702	.694
Orig. space	.813	.416	.265	.145	.138	.777	.756	.715	.534	–	–	–	–

Table 2. A subset of the RR values for individual cases of NMF subspace representations of the ORL face images selected from the set of all results used for the construction of the plots mentioned above.

explicit vector sparsity constraints introduced by Hoyer and simultaneously provides a truly parts-based vector basis of the NMF subspace. The main goal of the comparative computer experiments included in this study was to benchmark the results of the image object recognition with occlusions achieved by the above mentioned three NMF methods for a variety of recognition conditions. We decided to choose for these experiments the facial image data from public databases, since these data are freely available for other experimenters, and are most suitable for studying the algorithmic efficiency of the parts-based principle within the area of NMF data representation and reduction approaches to image object recognition under occlusions.

During the preparation of the extensive set of computer experiments, several methodological issues have revealed which had not been addressed in the existing papers. We have analyzed these issues and based on the results we modified the organization of our experiments. First of all, we have confirmed the published information about using the raw facial image data for benchmarking that suffers from the fact that, besides the relevant face-related pixels,

Subspace dimension	No occ.	Black occ.				Interpolated occ.				Masked occ.			
		Max	Mean	Min	LE+RE	Max	Mean	Min	LE+RE	Max	Mean	Min	LE+RE
Lee-Seung NMF – Original (top) / Cropped (middle) / Registered (bottom) images													
25	.926	.903	.861	.779	.727	.940	.922	.882	.867	.948	.940	.933	.941
250	.992	1.00	.986	.977	.941	.999	.99	.985	.977	1.00	.994	.985	.985
Orig. space	.956	.941	.909	.882	.772	.955	.941	.926	.911	–	–	–	–
25	.691	.635	.441	.316	.205	.695	.539	.391	.272	.754	.655	.551	.485
250	.808	.813	.683	.595	.338	.873	.788	.725	.595	.858	.798	.772	.625
Orig. space	.713	.698	.450	.279	.139	.757	.685	.588	.441	–	–	–	–
25	.720	.709	.493	.235	.139	.806	.65	.514	.439	.813	.704	.623	.588
250	.889	.739	.631	.507	.279	.910	.846	.822	.705	.903	.863	.808	.764
Orig. space	.802	.713	.452	.264	.176	.823	.746	.669	.485	–	–	–	–
Modified Hoyer NMF – Original (top) / Cropped (middle) / Registered (bottom) images													
25	.875	.844	.833	.814	.720	.903	.870	.830	.823	.911	.893	.875	.867
250	.933	.933	.905	.882	.808	.933	.917	.904	.875	.948	.939	.919	.933
Orig. space	.956	.941	.909	.882	.772	.955	.941	.926	.911	–	–	–	–
25	.463	.444	.310	.176	.169	.459	.388	.310	.264	.473	.430	.404	.338
250	.691	.665	.475	.345	.213	.777	.615	.502	.352	.784	.674	.588	.507
Orig. space	.713	.698	.450	.279	.139	.757	.685	.588	.441	–	–	–	–
25	.485	.465	.321	.176	.183	.495	.413	.330	.290	.503	.423	.382	.286
250	.801	.671	.439	.294	.191	.828	.716	.639	.462	.858	.769	.682	.698
Orig. space	.802	.713	.452	.264	.176	.823	.746	.669	.485	–	–	–	–
Modular NMF – Original (top) / Cropped (middle) / Registered (bottom) images													
25	.852	.873	.780	.660	.485	.851	.805	.748	.632	.851	.829	.807	.808
250	.963	.970	.959	.941	.889	.970	.954	.941	.919	.970	.959	.948	.933
Orig. space	.956	.941	.909	.882	.772	.955	.941	.926	.911	–	–	–	–
25	.683	.665	.491	.397	.183	.754	.638	.568	.316	.725	.650	.595	.507
250	.654	.665	.515	.404	.301	.725	.615	.514	.426	.710	.657	.622	.522
Orig. space	.713	.698	.45	.279	.139	.757	.685	.588	.441	–	–	–	–
25	.676	.628	.504	.411	.227	.710	.620	.570	.343	.717	.652	.610	.514
250	.698	.680	.563	.441	.264	.717	.646	.600	.483	.754	.696	.644	.602
Orig. space	.802	.713	.452	.264	.176	.823	.746	.669	.485	–	–	–	–

Table 3. A subset of the RR values for individual cases of NMF subspace representations of the YALE face images selected from the set of all results used for the construction of the plots mentioned above.

each image contains also the pixels from the background. These pixels can be in some sense even more informative than the facial pixels. The results of our experiments showed that classification of the face images significantly depends on the background presence. For ensuring a correct benchmark reference we proposed to crop all training and testing face images. Furthermore, for the sake of adapting the facial data for NMF parts-based representations, we proposed to normalize the positions of the explicit face parts (i.e., left eye, right eye, etc.) by the geometrical registration.

As for the issue of simulation of the partial object occlusion in images, besides usually used full (black) occlusions, we have introduced also interpolated and masked occlusions. The goal of the interpolated occlusions was to simulate a situation in which the intensities in occluded areas are being reconstructed by interpolation from nearest unoccluded facial pixels. The masked occlusions tried to simulate a situation in which the occluded image pixels were correctly excluded from the NMF calculations.

The detailed evaluation of the influence of various aspects on the Recognition Rates achieved by three NMF methods compared is given in the previous sections. The following general conclusions can be drawn on the basis of this evaluation. For the NMF benchmark studies, it is recommended to use cropped and registered facial image data. For recognition cases with full (nonsuppressed) occlusions and needing to maximally reduce dimension of the data representation, the Modular NMF method is recommended. For cases with other types of occlusions and without restriction on data dimension, the conventional Lee-Seung NMF algorithm slightly overcomes the two others. When the situation allows application of the masked approach to the NMF recognition, there is no apparent advantage of preferring particular one of the compared NMF methods.
It should be noted that the application of the interpolated and masked occlusion compensation method is completely dependent on the existence of an algorithm which is capable to identify the locations of the image pixels belonging to occlusions. Not to mention that the masking procedure is extremely computationally demanding, since it requires all original training data to be available during classification.

7. Acknowledgements

The research has been partially supported by the Slovak Grant Agency for Science VEGA (grant no. 2/0019/10).

8. References

Bajla, I. & Soukup, D. (2007). Non-negative matrix factorization – a study on influence of matrix sparseness and subspace distance metrics on image object recognition, *in* D. Fofi & F. Meriaudeau (eds), *SPIE*, Vol. 6356 of *Proc. 8th International Conference on Quality Control by Artificial Vision*, pp. (635614)–1 – (635614)–12.

Beymer, D. & Poggio, T. (1995). Face recognition from one example view, Proc. 5th ICCV'95, pp. 500–507.

Buciu, I. (2007). Learning sparse non-negative features for object recognition, Proc. 3rd IEEE Int. Conference on Intelligent Computer Communications and Processing, ICCP 2007, Romania, pp. 73–79.

Buciu, I., Nikolaidis, N. & Pitas, I. (2006). A comparative study of nmf, dnmf, and lnmf algorithms applied for face recognition, Proc. 2nd IEEE-EURASIP Int. Symposium on Control, Communications, and Signal Processing, ISCCSP 2006, Morocco.

Buciu, I. & Pitas, I. (2004). A new sparse image representation algorithm applied to facial expression recognition, Proc. IEEE Workshop on Machine Learning for Signal Processing, Sao Luis, Brasil, pp. 539–548.

Duda, R., Hart, P. & Stork, D. (2001). *Pattern classification*, John Wiley and Sons, Inc., New York.

Feng, T., Li, S., Shum, H. & Zhang, H. (2002). Local nonnegative matrix factorization as a visual representation, *Second Int. Conf. on Development and Learning*, Proc. ICDL '02.

Guillamet, D. & Vitrià, J. (2003). Evaluation of distance metrics for recognition based on non-negative matrix factorization, *Pattern Recognition Letters* 24: 1599–1605.

Heiler, M. & Schnörr, C. (2006). Learning sparse representations by non-negative matrix factorization and sequential cone programming, *Journal of Machine Learning Research* 7: 1385–1407.

Hoyer, P. (2002). Nonnegative sparse coding, *Neural Networks for Signal Processing XII*, Proc. IEEE Workshop on Neural Networks for Signal Processing, pp. 557–565.

Hoyer, P. (2004). Nonnegative matrix factorization with sparseness constraints, *Journal of Machine Learning Research* 5: 1457–1469.

Jolliffe, I. (2002). *Principal Component Analysis, Second Edition*, Springer-Verlag, Inc., New York.

Kanade, T., Cohn, J. & Tian, Y. (2000). Comprehensive database for facial expression analysis, Proc. IEEE Int. Conference on Face and Gesture Recognition, pp. 46–53.

Kim, J., Choi, J., Yi, J. & Turk, M. (2005). Effective representation using ICA for face recognition robust to local distortion and partial occlusion, *IEEE Trans. on PAMI* 27(12): 1977–1981.

Lee, D. & Seung, H. (1999). Learning the parts of objects by non-negative matrix factorization, *Nature* 401: 788–791.

Lee, D. & Seung, H. (2001). Algorithms for non-negative matrix factorization, *Advances in Neural Information Processing*, Proc. NIPS 2000.

Leonardis, A. & Bischof, H. (1994). Robust recognition using eigenimages, *Computer Vision and Image Understanding* 78(1): 99–118.

Li, S., Hou, X., Zhang, H. & Cheng, Q. (2001). Learning spatially localized, parts-based representation, *Computer Vision and Pattern Recognition*, Vol. 1 of *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR'01*, pp. I–207 – I–212.

Ling, H. & Okada, K. (2006). Diffusion distance for histogram comparison, Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR I, pp. 246–253.

Ling, H., Okada, K., Ponce, J., Berg, T. & Everingham, M. (2006). Dataset issues in object recognition, *Toward Category-Level Object Recognition*, Vol. LNCS 4170 of *Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR I*, pp. 29–48.

Liu, W. & Zheng, N. (2004). Non-negative matrix factorization based methods for object recognition, *Pattern Recognition Letters* 25: 893–897.

Liu, W., Zheng, N. & Lu, X. (2003). Nonnegative matrix factorization for visual coding, *2nd Int. Conf. on Development and Learning*, Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing ICASSP 2003.

Mel, B. (1997). Combining color, shape, and texture histogramming in neurally inspired approach to visual object recognition, *Neural Computation* 9(4): 777–804.

Murase, H. & Nayar, S. (1995). Visual learning and recognition of 3-D objects from appearance, *International Journal of Computer Vision* 14: 5–24.

Paatero, P. & Taper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values, *Environmetrics* 5: 111–126.

Pascual-Montano, A., Carazo, J., Kochi, K., Lehman, D. & Pascual-Marqui, R. (2006). Nonsmooth nonnegative matrix factorization (nsNMF), *IEEE Trans. on PAMI* 528(3): 403–415.

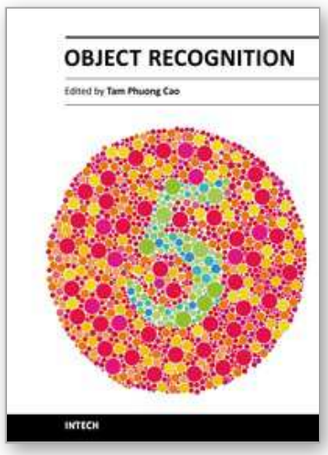
Shamir, L. (2008). Evaluation of face datasets as tools for assesing the performance of face recognition methods, *International Jornal of Computer Vision* 79: 225–230.

Soukup, D. & Bajla, I. (2008). Robust object recognition under partial occlusions using nmf, *Computational Intelligence and Neuroscience* 2008: ID 857453; DOI 10.1155/2008/857453.

Spratling, M. (2006). Learning image components for object recognition, *Journal of Machine Learning Research* 7: 793–815.

Turk, M. & Pentland, A. (1991). Eigenfaces for recognition, *Journal of Cognitive Neuroscience* 3(1): 71–86.

Yoshimura, S. & Kanade, T. (1994). Fast template matching based on the normalized correlation by using multiresolution eigenimages, Proc. IROS'94, pp. 2086–2093.



Object Recognition

Edited by Dr. Tam Phuong Cao

ISBN 978-953-307-222-7
Hard cover, 350 pages
Publisher InTech
Published online 01, April, 2011
Published in print edition April, 2011

Vision-based object recognition tasks are very familiar in our everyday activities, such as driving our car in the correct lane. We do these tasks effortlessly in real-time. In the last decades, with the advancement of computer technology, researchers and application developers are trying to mimic the humanâ€™s capability of visually recognising. Such capability will allow machine to free human from boring or dangerous jobs.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Ivan Bajla, Daniel Soukup and Svorad Štolc (2011). Occluded Image Object Recognition using Localized Nonnegative Matrix Factorization Methods, Object Recognition, Dr. Tam Phuong Cao (Ed.), ISBN: 978-953-307-222-7, InTech, Available from: http://www.intechopen.com/books/object-recognition/occluded-image-object-recognition-using-localized-nonnegative-matrix-factorization-methods

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen