# MULTIPLE NONNEGATIVE-MATRIX FACTORIZATION OF DYNAMIC PET IMAGES

*Jong-Hoon Ahn*[†], *Sang-Ki Kim*[§], *Jong-Hoon Oh*[†], *Seungjin Choi*[§]

[†]Department of Physics, POSTECH, Korea
[§]Department of Computer Science, POSTECH, Korea
{*jonghun, ihistory, jhoh, seungjin*}*@postech.ac.kr*

## ABSTRACT

We propose an extension of nonnegative matrix factorization (NMF) to multilayer network model for dynamic myocardial PET image analysis. NMF has been previously applied to the analysis and shown to successfully extract three cardiac components and time-activity curve from the image sequences. Here we apply triple nonnegative-matrix factorization to the dynamic PET images of dog and show details of cardiac components. We think of the multiple nonnegative-matrix factorization as a model that can learn a hierarchical-parts representation.

## 1. INTRODUCTION

Positron emission tomography (PET) is a technique for measuring the concentrations of positron-emitting radio-isotopes within the tissue of living subjects. One of the most important functions of PET is its ability to model biological and physiological functions in the body by detection and modeling of regional concentrations of radioactivity in a particular organ. It is then useful and important to obtain quantitative information on the regional myocardial blood flow using dynamic PET scan.

Since the pioneering work by Barber on the quantitative analysis [1] through principal components, several methods and algorithms such as factor analysis have been applied to analyze the dynamic PET image sequences. It is recently shown that non-negative matrix factorization (NMF) could be a suitable algorithm for the analysis. Most approaches basically assume Gaussian statistics, which may not be appropriate for gamma camera images, whereas NMF utilize Poisson statistics as a noise model. This point fits in with the fact that the gamma camera images really represent some sort of photon counts.

However, the considerable amount of statistical noise generated in short dynamic frames and small number of image sequences make it difficult to identify the cardiac components. The NMF algorithm is practically deterministic model, and does not take the generated noise into account. In order to describe and get rid of noisy factor, a probabilistic model, which can originally give NMF in zero noise level, is required. But it seems so difficult or intractable even if it is really derived. We also point out that the successful segmentation of cardiac components actually comes from the constraint of non-negativity, not from using the cost function associated to poisson noise model. Because of these factors, when we attempt to obtain details of cardiac components more than three, NMF do not show them. If the number of hidden variables is more than three, each part is broken down and its time-activity curve is shaken.

In this paper, we propose a hierarchical extension of NMF for extracting smaller parts than those extracted by NMF. We consider a multi-layer generative network model with all weights and hidden variables nonnegative. A multiplicative algorithm for training the network model, which is guaranteed to converge monotonically without the need for setting any adjustable parameters such as learning rate, is also proposed. This model parallels a multiple nonnegative-matrix factorization of a given nonnegative matrix. We treat the method as a model that can learn hierarchical parts representation from complex structure data such as face and hands. We show the possibility that the new method may be also suitable for dynamic PET image analysis in nuclear medicine and report it. [2] To study the application to dynamic PET image analysis, the method was applied to myocardial $H_2^{15}O$ PET images, in which we attempt to segment detailed cardiac components and derive their time-activity curves. The results are fully discussed.

## 2. NONNEGATIVE MATRIX FACTORIZATION

Let a set of $N$ training images be given as an $p \times N$ matrix $\boldsymbol{V}$, with each column consisting of the $p$ non-negative pixel values of an image. Denote a set of $q \leq p$ basis images by a $p \times q$ matrix $\boldsymbol{W}$. Each image can be represented as a linear combination of the basis images using the approximate factorization

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

where $\boldsymbol{H} \in \boldsymbol{R}^{q \times N}$ is the encoding variable matrix. Dimension reduction is achieved when $q < p$.

The PCA factorization requires that the basis images (columns of $\boldsymbol{W}$ be orthonormal and the rows of $\boldsymbol{H}$ be mutually orthogonal. It imposes no other constraints than the orthogonality and hence allows the entries of $\boldsymbol{W}$ and $\boldsymbol{H}$ to be of arbitrary sign. Many basis images, or eigenfaces in the case of face recognition, lack intuitive meaning, and a linear combination of the bases generally involves complex cancellations between positive and negative numbers. The NMF representations allow only positive coefficients and thus non-subtractive combinations. [3] [4] [5]

NMF imposes the non-negativity constraints instead of the orthgonality. As the consequence, the entries of $\boldsymbol{W}$ and $\boldsymbol{H}$ are all non-negative, and hence only non-subtractive combinations are allowed. This is believed to be compatible to the intuitive notion of combining parts to form a whole, and is how NMF learns a parts-based representation. It is also consistent with the physiological facts that the firing rates are non-negative and the signs of synapses do not change.

NMF uses the I-divergence of $\boldsymbol{V}$ from $\boldsymbol{WH}$, defined as

$$\begin{aligned} & D(\boldsymbol{V}||\boldsymbol{WH}) \\ = & \sum_{i,j}\left(\boldsymbol{V}_{ij}\log\frac{\boldsymbol{V}_{ij}}{(\boldsymbol{WH})_{ij}} - \boldsymbol{V}_{ij} + (\boldsymbol{WH})_{ij}\right) \end{aligned} \quad (2)$$

as the measure of fitness for factorizing $\boldsymbol{V}$ into $\boldsymbol{WH}$. An NMF factorization is defined as

$$\min_{\boldsymbol{W},\boldsymbol{H}} \quad D(\boldsymbol{V}||\boldsymbol{WH}) \quad (3)$$

$$s.t \quad \boldsymbol{W},\boldsymbol{H} \geq 0 \quad (4)$$

where $\boldsymbol{W}, \boldsymbol{H} \geq \boldsymbol{O}$ means that all entries of $\boldsymbol{W}$ and $\boldsymbol{H}$ are non-negative. $D(\boldsymbol{V}||\boldsymbol{WH})$ reduces to Kullback-Leibler divergence when $\sum_{i,j}\boldsymbol{V}_{ij} = \sum_{i,j}(\boldsymbol{WH})_{ij} = 1$. The above optimization can be done by using multiplicative update rules.

$$\begin{aligned} \boldsymbol{H}_{a\mu} &\leftarrow \boldsymbol{H}_{a\mu}\frac{\sum_i \boldsymbol{W}_{ia}\boldsymbol{V}_{i\mu}/(\boldsymbol{WH})_{i\mu}}{\sum_k \boldsymbol{W}_{ka}} \\ \boldsymbol{W}_{ia} &\leftarrow \boldsymbol{W}_{ia}\frac{\sum_\mu \boldsymbol{H}_{a\mu}\boldsymbol{V}_{i\mu}/(\boldsymbol{WH})_{i\mu}}{\sum_\nu \boldsymbol{H}_{a\nu}} \end{aligned} \quad (5)$$

The algorithm performs both learning and inference simultaneously. That is, it both learns a set of basis images and infers values for the hidden variables from the visible variables. Although the generative model is linear, the inference computation is nonlinear due to the non-negativity constraints. The computation is similar to maximum likelihood reconstruction in emission tomography, and deconvolution of blurred astronomical images.

Although NMF is successful in learning facial parts and semantic topics, this success does not imply that the method can learn parts from any database, such as images of objects viewed from extremely different viewpoints, or highly articulated objects. Learning parts for these complex cases is likey to require fully hierarchical models with multiple levels of hidden variables, instead of the single level in NMF. [3]

## 3. MULTIPLE NONNEGATIVE-MATRIX FACTORIZATION

There are a large number of neurons in inferior temporal cortex of monkeys which seem to encode an overall shape of biologically important objects - not specific features or parts. [6] The finding agrees with hierarchical theories of object perception. According to these theories, cells in the cortical areas code elementary features such as line orientation and color. The outputs from these cells are then combined by detectors sensitive to higher-order features such as corners or intersections, an idea consistent with the findings of Hubel and Wiesel. The process is continued as each successive stage codes more complex combinations. At the top of the chain are IT neurons, selective for complex shapes like hands or faces. A huge number of hierarchical models for object recognition have been proposed over the years. Some of them were inspired by the desire to build intelligent machines, others by the desire to describe human recognition processes. [7]

In this paper, we will try to invert the hierarchical recognition processes and view the visual perception as a hypothesis testing process. Helmholtz, in his doctrine of unconscious inference, argued that perceptions are formed by the interaction of bottom-up sensory data with top-down expectations. According to one interpretation of this doctrine, perception is a procedure of sequential hypothesis testing. We propose a new algorithm, called multiple nonnegative-matrix factorization, that realizes this interpretation in layered networks. It uses top-down connections to generate hypotheses, and bottom-up connections to revise them.

How can we build such a hierarchical structure of special neurons which are responsible only for local sensory features by multi-layer generative model? We have already known that NMF can give local features and sparse codes from given non-negative data set. It is just a single-layer generative network and gives local features and sparse codes. Then what about multi-layer network with nonnegative weights and hidden variables? It is what we will introduce in this section. The previous notations of $\boldsymbol{W}$ and $\boldsymbol{H}$ are changed into $\boldsymbol{W}^{(1)}$ and $\boldsymbol{H}^{(1)}$. We start by assuming further generations:

$$\boldsymbol{H}^{(l)} = \boldsymbol{W}^{(l+1)}\boldsymbol{H}^{(l+1)} \quad (6)$$

where $l = 1, 2, \cdots, L-1$. Then we can construct a $L$-layer nonnegative networks as shown in fig. 1. It is basically a generative model and constructs the nonnegative input data at its bottom layer. Before we show how it finds the hierarchical features from the nonnegative data, we should think

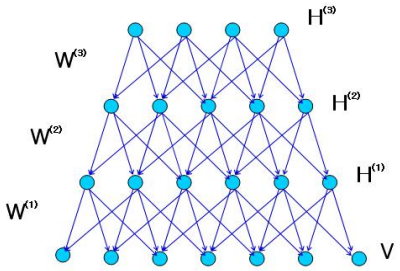


**Fig. 1**. Multilayer network model of multiple nonnegative-matrix factorization: This shows an architecture of the multilayer network with an input layer and three hidden layers.

of the algorithm which we can find the optimized values of weights and hidden variables with.

In order to extend NMF to multiple nonnegative-matrix factorization, let us intruduce two matrices $\boldsymbol{R}$ and $\boldsymbol{N}$ to the update rules of equation (5):

$$\begin{aligned}
\boldsymbol{H}_{a\mu} &\leftarrow \boldsymbol{H}_{a\mu}\frac{\sum_i \boldsymbol{W}_{ia}\frac{\boldsymbol{V}_{i\mu}}{(\boldsymbol{WH})_{i\mu}}}{\sum_k \boldsymbol{W}_{ka}} \\
&= \boldsymbol{H}_{a\mu}\frac{\sum_i \boldsymbol{W}_{ia}\boldsymbol{R}_{i\mu}}{\sum_k \boldsymbol{W}_{ka}\boldsymbol{N}_{k\mu}} \\
&= \boldsymbol{H}_{a\mu}\frac{(\boldsymbol{W}^T\boldsymbol{R})_{a\mu}}{(\boldsymbol{W}^T\boldsymbol{N})_{a\mu}} \\
\boldsymbol{W}_{ia} &\leftarrow \boldsymbol{W}_{ia}\frac{\sum_\mu \frac{\boldsymbol{V}_{i\mu}}{(\boldsymbol{WH})_{i\mu}}\boldsymbol{H}_{a\mu}}{\sum_\nu \boldsymbol{H}_{a\nu}} \\
&= \boldsymbol{W}_{ia}\frac{\sum_\mu \boldsymbol{R}_{i\mu}\boldsymbol{H}_{a\mu}}{\sum_\nu \boldsymbol{N}_{i\nu}\boldsymbol{H}_{a\nu}} \\
&= \boldsymbol{W}_{ia}\frac{(\boldsymbol{RH}^T)_{ia}}{(\boldsymbol{NH}^T)_{ia}} \qquad (7)
\end{aligned}$$

where $\boldsymbol{R} = \boldsymbol{R}^{(1)}$ is a matrix of ratios of $\boldsymbol{V}$ to $\boldsymbol{WH}$ and $\boldsymbol{N} = \boldsymbol{N}^{(1)}$ is a normalizer with all elements 1. If we put $\boldsymbol{R} = \boldsymbol{V}$ and $\boldsymbol{N} = \boldsymbol{WH}$, the above rules are for least squared error under nonnegativity. In fact the formulation includes all NMF for optimizing any form of cost function. If we also obtain the matrix $\boldsymbol{R}^{(l)}$ and $\boldsymbol{N}^{(l)}$ at $l$th layer of the networks, we can optimize the multi-layer networks by using the same update rules at all layers:

$$\boldsymbol{W}^{(l)}_{ia} \leftarrow \boldsymbol{W}^{(l)}_{ia}\frac{(\boldsymbol{R}^{(l)}\boldsymbol{H}^{(l)T})_{ia}}{(\boldsymbol{N}^{(l)}\boldsymbol{H}^{(l)T})_{ia}} \qquad (8)$$

and

$$\boldsymbol{H}^{(L)}_{a\mu} \leftarrow \boldsymbol{H}^{(L)}_{a\mu}\frac{(\boldsymbol{W}^{(L)T}\boldsymbol{R}^{(L)})_{a\mu}}{(\boldsymbol{W}^{(L)T}\boldsymbol{N}^{(L)})_{a\mu}} \qquad (9)$$

Fortunately, the $\boldsymbol{R}^{(l)}$ and $\boldsymbol{N}^{(l)}$ matrices are calculated by the following up-propagation rules:

$$\begin{aligned}
\boldsymbol{R}^{(l+1)} &= \boldsymbol{W}^{(l)T}\boldsymbol{R}^{(l)} \\
\boldsymbol{N}^{(l+1)} &= \boldsymbol{W}^{(l)T}\boldsymbol{N}^{(l)} \qquad (10)
\end{aligned}$$

where $l = 1, 2, \cdots, L-1$.

However, the multilayer network model without using nonlinear transfer function could be functionally degenerated into a single layer model. The data reconstructed from all $\boldsymbol{W}^{(i)}$ and $\boldsymbol{H}^{(L)}$ just fill a linear subspace, and cannot make a nonlinear manifold which the given data matrix $\boldsymbol{V}$ orginally lies on. Thus we should consider the nonnegative multi-layer model with nonlinear transfer functions instead of equation (6):

$$\boldsymbol{H}^{(l)}_{a\mu} = \mathrm{g}\left((\boldsymbol{W}^{(l+1)}\boldsymbol{H}^{(l+1)})_{a\mu}\right) \qquad (11)$$

where the nonlinear transfer function g must output a nonnegative value. Although we insert the nonlinearity between successive layers, the NMF algorithms (8) and (9) are kept on. The up-propagation rules (10) are then changed into

$$\begin{aligned}
\boldsymbol{R}^{(l+1)}_{i\mu} &= (\boldsymbol{W}^{(l)T}\boldsymbol{R}^{(l)})_{i\mu}\mathrm{g}'\left((\boldsymbol{W}^{(l+1)}\boldsymbol{H}^{(l+1)})_{i\mu}\right) \\
\boldsymbol{N}^{(l+1)}_{i\mu} &= (\boldsymbol{W}^{(l)T}\boldsymbol{N}^{(l)})_{i\mu}\mathrm{g}'\left((\boldsymbol{W}^{(l+1)}\boldsymbol{H}^{(l+1)})_{i\mu}\right)
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{R}^{(1)}_{i\mu} &= \frac{\boldsymbol{V}_{i\mu}}{(\boldsymbol{W}^{(1)}\boldsymbol{H}^{(1)})_{i\mu}}\mathrm{g}'\left((\boldsymbol{W}^{(1)}\boldsymbol{H}^{(1)})_{i\mu}\right) \\
\boldsymbol{N}^{(1)}_{i\mu} &= \mathrm{g}'\left((\boldsymbol{W}^{(1)}\boldsymbol{H}^{(1)})_{i\mu}\right). \qquad (12)
\end{aligned}$$

The proposed algorithm is similar with error up-propagation algorithm for multi-layer neural network model, but is also different in several aspects. [8] The error up-propation algorithm propagates the error, whereas the proposed algorithm should propagate normalizing factors as well as ratios of inputs to reconstructed values. Besides, note that the ratios and normalizers should be up-propagated every time before each weights matrix $\boldsymbol{W}^{(l)}$ is updated.

## 4. APPLICATION TO DYNAMIC PET IMAGES

We analyzed $H_5^{15}O$ PET scans performed on seven dogs at rest and after pharmacological stress using Adenosine or Dipyridamole [2] [9]. All the scans were acquired with an ECAT EXACT 47 scanner (Simens CTI,Knoxville, USA) wich has an intrinsic resolution of 5.2 nn full width at half maximum and images 47 continuous planes with thickness of 3.4 mm simultaneously for a longitudinal field of view of 16.2 cm. Before $H_5^{15}O$ administration, transmission scanning was performed using three Ge-68 rod sources for attenuation correction. Dynamic emission scans (5sec$\times$12,

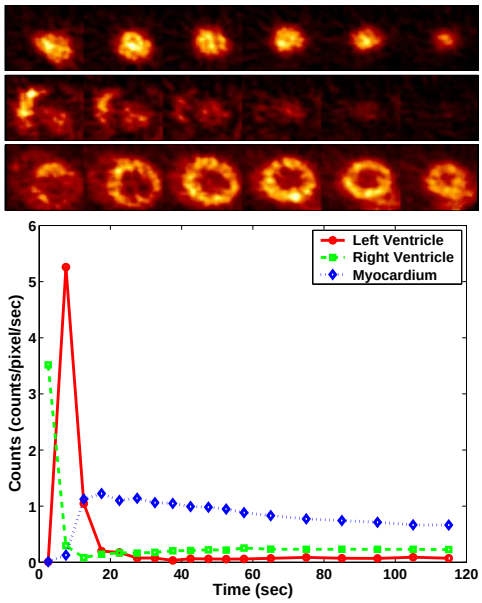


**Fig. 2**. Basis images (column vectors of $\boldsymbol{W}^{(1)}\boldsymbol{W}^{(2)}$) and their time-activity curves (row vectors of $\boldsymbol{H}^{(2)}$ obtained using triple factorization (two layer networks with nonnegativity) from a dog PET data at rest

10sec×9, 30 sec×3) were initiated simultaneously with the injection of 555-740 MBq $H_5^{15}O$. Transaxial images were reconstructed by means of a filtered back-projection algorithms as 128 ×123×47 matrices with a size of 2.1×2.1×3.4 mm.

The initial eighteen frames (two minutes) of PET images were used for analysis. The dynamic PET images were re-oriented to short axis and were re-sampled to produce 1-cm-thick slices in order to increase the signal to noise ratio. Only the cardiac regions were masked to remove extra cardiac components and to reduce the quantity of data and hence the burden of computation. The resulting masked images with dimension of 32×32×6×18 (pixel × pixel × plane × frame) were reformulated to 6144×18 (pixel × frame) data matrix $\boldsymbol{X}$.

The fig. 2 shows the basis images and their time-activity curve extracted by using the two layer nonnegative network. The three bases of the figure are left ventricle, right ventricle and myocardium, respectively. The bases represent the column vectors of the product of $\boldsymbol{W}^{(1)}$ and $\boldsymbol{W}^{(2)}$ and the values of 18 points of curves are from row vectors of $\boldsymbol{H}^{(2)}$. These results are converged to those from NMF, as the number of hidden variables of the bottom layer increases. 12 hidden variables are used in this analysis. Thus, the 12 detailed parts are extrated as shown in fig. 3. The three cardiac components are constructed by weighted summation of 12 parts. The weight values are indicated in fig. 5.It seems that a few of parts contain redundant information for three cardiac components. But we can learn that the detailed parts function as independent factors by carefully considering the weight values of fig. 5. For example, the 3, 4, 5, 6 and 10th parts look like the intensity-scaled replica of left ventricle by the curves of fig. 4. But the detailed parts are also used in constructing the other cardiac components with different weights. This tells that the detailed parts contribute to constructing the three major features independently. The right ventricle consists of 2, 7, 8, 9 and 12th parts of figure 4. We can discriminate 2, 7 and 12th parts from 8th and 9th parts without trouble. The former parts don't contribute to left ventricle, but the latter parts considerably contribute to right ventricle. The myocardium is mainly from 11th part and partly from the first part. The myocardium also uses the 5th part fairly, which the left ventricle shares.

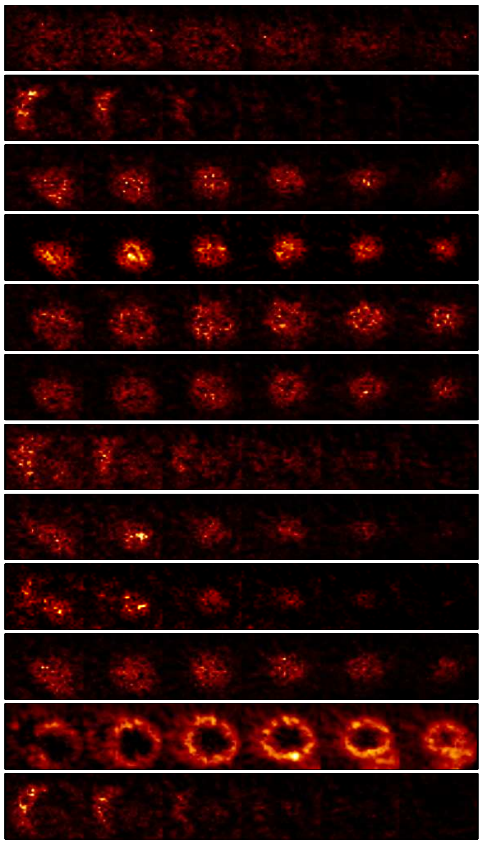

**Fig. 3**. Basis images (column vectors of $\boldsymbol{W}^{(1)}$) obtained using triple factorization from a dog PET data at rest

## 5. ACKNOWLEDGMENT

This work was supported by Korea Ministry of Science and Technology under Brain Science and Engineering Research Program, KOSEF 2000-2-20500-009-5, and Brain Korea 21 in POSTECH.

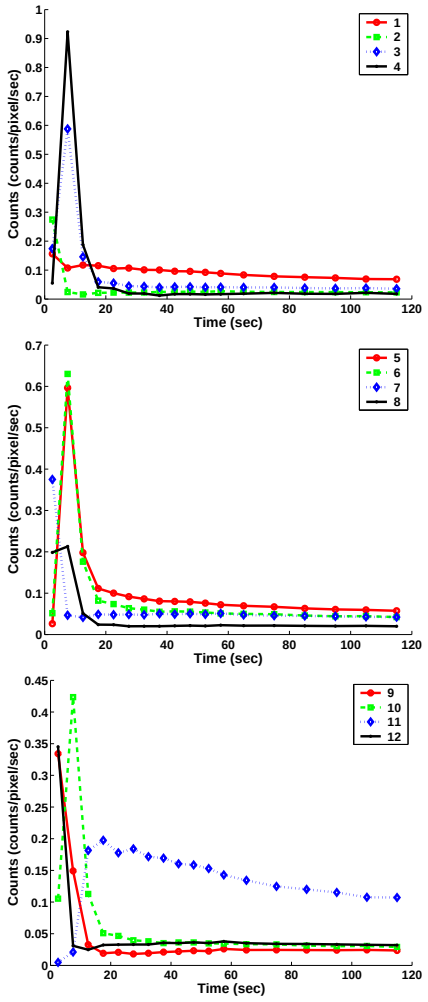


**Fig. 4**. Time-activity curves (row vectors of $\boldsymbol{W}^{(2)}\boldsymbol{H}^{(2)}$) obtained using triple factorization from the dog PET data at rest

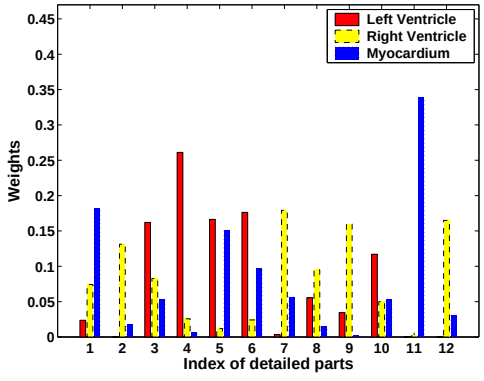


**Fig. 5**. Normalized coefficients of detailed parts used in contructing the three cardiac components

## 6. REFERENCES


[1] D. C. Barber, "The use of principle components in the quantitative analysis of gamma camera dynamic studies," *Physiology and Medical Biology*, vol. 25, pp. 283–292, 1980.

[2] J. S. Lee, D. D. Lee, S. Choi, K. S. Park, and D. S. Lee, "Nonnegative matrix factorization of dynamic images in nuclear medicine," in *IEEE Medical Imaging Conference*, 2001.

[3] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, Oct. 1999.

[4] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems*, 2001, vol. 13.

[5] S. Z. Li, X. W. Hou, H. J. Zhang, and Q. S. Cheng, "Learning spatially localized, parts-based representation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Kauai, Hawaii, 2001, pp. 207–212.

[6] D. J. Perrett and M. W. Oram, "Neurophysiology of shape processing," *Image and Vision Computing*, vol. 11, pp. 317–333, 1993.

[7] M. S. Gazzaniga, R. B. Ivry, and G. R. Mangum, *Cognitive Neuroscience: The Biology of the Mind*, W. W. Norton & Company, New York, 2001.

[8] J. H. Oh and H. S. Seung, "Learning generative models with the up-propagation algorithm," in *Advances in Neural Information Processing Systems*. 1997, pp. 605–611, MIT.

[9] J. S. Lee, D. D. Lee, S. Choi, and D. S. Lee, "Application of non-negative matrix factorization to dynamic positron emission tomography," in *Proc. ICA'01*, 2001, pp. 629–632.