Lifting Wavelet Based Cognitive Vision System

Yuka Higashijima, Shigeru Takano, and Koichi Niijima

Department of Informatics, Kyushu University
744, Motooka, Nishi-ku, Fukuoka 819-0395, Japan
{y-higasi,takano,niijima}@i.kyushu-u.ac.jp

Abstract. This paper presents a cognitive vision system based on the learning of lifting wavelets. The learning process consists of four steps: 1. Extract training and query object images automatically from adjacent video frames using our proposed cosine-maximization method; 2. Compute autocorrelation vectors from the extracted training images, and their discriminant vectors by linear discriminant analysis; 3. Map the autocorrelation vectors onto the discriminant vector space to obtain feature vectors; 4. Learn lifting parameters in the feature vectors using the idea of discriminant analysis. The recognition of a query object is performed by measuring cosine distance between its feature vector and the feature vectors for training object images. Our experimental results on vehicle types recognition show that the proposed system performs better than the discriminant analysis of original images.

1 Introduction

With recent advances of network cameras, their practical use in monitoring is rapidly broadening. The network camera can capture in real-time a large amount of objects such as vehicles and walkers, moving on streets and intersections. The images of these objects are memorized in its server after compression. By monitoring the stored images, we can recognize, for example, traffic conditions in the streets and intersections. However, present network cameras are not capable of recognizing vehicle types, counting the number of vehicles moving for a specific period of time, or detecting the vehicles of illegal parking. For the network cameras to possess such kinds of functions, more intelligent cognitive vision systems are needed.

There are many approaches for detecting and recognizing moving vehicles. Schneiderman et al. [3] studied a statistical model for vehicle detection. Another different statistical model was presented by Weber et al. [10]. They extracted a set of local features from each vehicle image and applied the EM algorithm to learn the parameters contained in the probability distribution of the set. Papageorgiou et al. [2] extracted the features of objects using Haar wavelet transform, and classified the objects by applying Support Vector Machines (SVMs) to the features. Also, Haar wavelets and SVMs were combined for vehicle detection

Proceedings of the ICVS International Cognitive Vision Workshop – ICVW 2007
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany

by Sun et al. [4]. They have proposed three feature selection schemes based on Haar wavelets for detecting moving vehicles by the use of SVMs. See e.g. [5] for a recent review on vision-based on-road vehicle detection systems.

In contrast to these approaches, our cognitive vision system is based on the learning of lifting parameters contained in the lifting filters. Training object images are automatically extracted as follows: We compute the difference images between adjacent video frames containing various objects as well as the target objects. To extract only the target objects, we learn a lifting wavelet filter at their centres using our cosine-maximization method [6–9]. By applying the learned filter to another set of difference images, we extract object regions around the detected centres after the dilation operation. The object regions are mapped into the original video frame, and rectangular image regions containing the objects are cut off from the video frame. We put the values of pixels 0 in the rectangular region except for the objects. These rectangular images are used as training images.

From the training images, we compute three-order autocorrelation vectors. Linear discriminant analysis is applied to the vectors to compute the discriminant vectors. Next, we apply the dyadic wavelet transform to the training images to get low-pass and high-pass images. Lifted high-pass images are constructed from these images. By combining the lifted images with the discriminant vectors, we construct feature vectors for the training images. We learn lifting parameters in the feature vectors such that the Fisher criterion is maximized.

The recognition of a query object is performed as follows: We extract a rectangular image containing moving objects as for training images. The dyadic wavelet transform is applied to the rectangular image for obtaining low-pass and high-pass images. By combining these images with the learned parameters, we construct several lifted high-pass images. Furthermore, we compute a feature vector by the combination of these high-pass images with the discriminant vectors. To recognize the query object, the query feature vector is compared with the feature vectors for the training set by the use of cosine distance.

In experiments, we address the problem of detection and recognition of vehicles moving on a street in our campus. These images can always be captured by the web camera equipped at the window of our laboratory. From the captured video sequences, the images of vehicles such as sedan, taxi, van, truck, and bus, are extracted using our technique. The recognition of vehicle types is carried out based on the proposed recognition method.

The remainder of this paper is structured as follows. Section 2 introduces the lifting dyadic wavelets and autocorrelation vectors. In Section 3, we present an algorithm for learning lifting parameters. Section 4 is devoted to our recognition system. Section 5 shows the recognition results of vehicle types. Finally, we conclude with Section 6.

Proceedings of the ICVS International Cognitive Vision Workshop – ICVW 2007
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

2 Lifting Dyadic Wavelets and Autocorrelation Vectors

2.1 Lifting Dyadic Wavelet Transform

Let $\{h_n^o, g_n^o, \tilde{h}_n^o, \tilde{g}_n^o\}$ be an initial set of dyadic wavelet filters. The h_n^o and g_n^o are called low-pass and high-pass analysis filters, respectively, and the $\tilde{h}_n^o$ and $\tilde{g}_n^o$ are low-pass and high-pass synthesis filters, respectively. A lifting scheme for dyadic wavelet is given by

$$\begin{aligned} h_n &= h_n^o, \\ g_n &= g_n^o - \sum_l \lambda_l h_{n-l}^o, \\ \tilde{h}_n &= \tilde{h}_n^o + \sum_l \lambda_{-l} \tilde{g}_{n-l}^o, \qquad \tilde{g}_n = \tilde{g}_n^o. \end{aligned} \tag{1}$$

Here λ_l's are called the lifting parameters.

The lifting wavelet coefficients for an image $u_{i,j}$ are computed as follows: An application of h_n^o in vertical direction to $u_{i,j}$ yields

$$C_{m,k}^{col} = \sum_j h_j^o u_{m,k+j}.$$

By applying the lifting filter (1) in horizontal direction to $C_{m,k}^{col}$, we get the lifting wavelet coefficients

$$D_{m,k} = \sum_i g_i^d C_{m+i,k}^{col}. \tag{2}$$

Here g_i^d's are given by

$$g_i^d = g_i^o - \sum_{l=-L}^{L} \lambda_l^d h_{i-l}^o, \qquad i = -K-L, ..., K+L+1,$$

where λ_l^d's represent lifting parameters in horizontal direction. We assume that the index i of the filter h_i^o ranges from $-K$ to $K+1$.

Next, we compute

$$C_{m,k}^{row} = \sum_i h_i^o u_{m+i,k}$$

by the application of h_n^o in horizontal direction. Applying the lifting filter (1) in vertical direction to $C_{m,k}^{row}$ gives

$$E_{m,k} = \sum_j g_j^e C_{m,k+j}^{row}. \tag{3}$$

Here g_j^e's denote

$$g_j^e = g_j^o - \sum_{l=-L}^{L} \lambda_l^e h_{j-l}^o, \qquad j = -K-L, ..., K+L+1,$$

where λ_l^e's represent lifting parameters in vertical direction.

It follows from (2) and (3) that

$$\begin{aligned} F_{m,k} &= D_{m,k} + E_{m,k} \\ &= F^o_{m,k} - \left(\sum_{l=-L}^{L} \lambda_l^d C_{m+l,k} + \sum_{l=-L}^{L} \lambda_l^e C_{m,k+l} \right), \end{aligned} \tag{4}$$

where

$$F^o_{m,k} = \sum_i g_i^o C^{col}_{m+i,k} + \sum_j g_j^o C^{row}_{m,k+j}, \qquad C_{m,k} = \sum_{i,j} h_i^o h_j^o u_{m+i,k+j}. \tag{5}$$

If the primal high-pass filter coefficients g_n^o are chosen as $g_0^o = g_2^o = -0.25\sqrt{2}$, $g_1^o = 0.5\sqrt{2}$ and $g_i^o = 0$ otherwise, then the lifting wavelet filter defined by (4) approximates an elliptic-type partial differential operator $Q(\lambda^d, \lambda^e)$ defined by

$$Q(\lambda^d, \lambda^e)u = -\left(\frac{\partial^2}{\partial s^2}(I_t u) + \frac{\partial^2}{\partial t^2}(I_s u) \right) - I(\lambda^d, \lambda^e)u. \tag{6}$$

Here, $u = u(s,t)$ is the continuous version of $u_{i,j}$, $I_t u$ and $I_s u$ represent the integral versions of $C^{col}_{m,k}$ and $C^{row}_{m,k}$ respectively, and $I(\lambda^d, \lambda^e)u$ corresponds to the last term of (4). The λ^d and λ^e are denoted by

$$\lambda^d = (\lambda^d_{-L}, ..., \lambda^d_L), \qquad \lambda^e = (\lambda^e_{-L}, ..., \lambda^e_L).$$

2.2 Autocorrelation Vectors

Feature vectors for object detection and recognition are preferred to be shift-invariant and additive. In this paper, we use autocorrelation vectors, which are shift-invariant and additive [1]. A three-order autocorrelation vector is constructed from the original image $u_{m,k}$ using three kinds of autocorrelation functions

$$x^o = \sum_{m=0}^{n_s-1} \sum_{k=0}^{n_t-1} u_{m,k}, \tag{7}$$

$$x^o((i_1, j_1)) = \sum_{m=0}^{n_s-1} \sum_{k=0}^{n_t-1} u_{m,k} u_{m+i_1,k+j_1}, \tag{8}$$

$$x^o((i_1, j_1), (i_2, j_2)) = \sum_{m=0}^{n_s-1} \sum_{k=0}^{n_t-1} u_{m,k} u_{m+i_1,k+j_1} u_{m+i_2,k+j_2}. \tag{9}$$

Here, (i_1, j_1) and (i_2, j_2) denote displacements, and their range is restricted to within a local 3×3 window. If we eliminate the displacements which are equivalent by the shift, the number of displacement patterns is 35. By arranging these functions, we construct a 35-dimensional autocorrelation vector

$$x^o = (x_1^o, x_2^o, ..., x_{35}^o)', \tag{10}$$

where $'$ indicates transpose. Similarly, we can define autocorrelation functions $x_i,\ i = 1, ..., 35$ for the lifted high-pass image $F_{m,k}$ as in (7), (8) and (9), and construct an autocorrelation vector as

$$x = (x_1, x_2, ..., x_{35})'. \tag{11}$$

3 Learning Algorithm

3.1 Discriminant Analysis

For convenience, we put $N = 35$. Assume that we have $M(\leq N)$ classes of training images, whose ν-th class consists of T images. Let us denote the τ-th training image in the ν-th class by $u_{i,j}^{\nu,\tau}$. From $u_{i,j}^{\nu,\tau}$, we compute the autocorrelation vectors $x^{o,\nu,\tau}$, which has the form (10).

We put $M_1 = M - 1$. The discriminant analysis is to compute the $N \times M_1$ discriminant matrix $A = (a^1\ a^2 ...\ a^{M_1})$ by solving the matrix equations

$$\Sigma_B^o A = \Sigma_W^o A \Lambda,$$
$$A' \Sigma_W^o A = E.$$

Here, Λ is a diagonal matrix, E is the unit matrix, and Σ_W^o and Σ_B^o denote the within-class and between-class covariance matrices represented, respectively, by

$$\Sigma_W^o = \frac{1}{MT} \sum_{\nu=1}^{M} \sum_{\tau=1}^{T} (x^{o,\nu,\tau} - \bar{x}^{o,\nu})(x^{o,\nu,\tau} - \bar{x}^{o,\nu})',$$
$$\Sigma_B^o = \frac{1}{M} \sum_{\nu=1}^{M} (\bar{x}^{o,\nu} - \bar{x}^{o})(\bar{x}^{o,\nu} - \bar{x}^{o})',$$

where $\bar{x}^{o,\nu}$ is the average of $x^{o,\nu,\tau}$ in the ν-th class, and $\bar{x}^o$ the average of whole vectors $x^{o,\nu,\tau}$.

3.2 Learning of Lifting Parameters

The autocorrelation vectors $x^{o,\nu,\tau}$ can be expanded using the discriminant vectors a^p as

$$x^{o,\nu,\tau} = \sum_{p=1}^{M_1} (a^p \cdot x^{o,\nu,t}) a^p,$$

because the remaining discriminant vectors $a^p,\ p = M, ..., N$ span the null-space. Here the symbol $\cdot$ indicates inner product.

Now, from the lifted high-pass image $F_{m,k}^{\nu,\tau}$ having the expression (4), we construct M_1 autocorrelation vectors $x^{\nu,\tau,p},\ p = 1, ..., M_1$ with the form (11). Using $x^{\nu,\tau,p}$, we define a vector

$$y^{\nu,\tau} = (a^1 \cdot x^{\nu,\tau,1},\ a^2 \cdot x^{\nu,\tau,2}, \ldots,\ a^{M_1} \cdot x^{\nu,\tau,M_1})', \tag{12}$$

which is a generalization of the feature vector

$$(a^1 \cdot x^{o,\nu,\tau}, a^2 \cdot x^{o,\nu,\tau}, \ldots, a^{M_1} \cdot x^{o,\nu,\tau})'.$$

We determine the lifting parameters contained in (12) component-wise. The within-class and between-class variances of the p-th components $y_p^{\nu,\tau}$ are given by

$$\rho_p = \frac{1}{MT}\sum_{\nu=1}^{M}\sum_{\tau=1}^{T}(y_p^{\nu,\tau} - \bar{y}_p^{\nu})^2, \qquad \sigma_p^2 = \frac{1}{M}\sum_{\nu=1}^{M}(\bar{y}_p^{\nu} - \bar{y}_p)^2, \tag{13}$$

respectively, where $\bar{y}_p^{\nu} = \sum_{\tau=1}^{T} y_p^{\nu,\tau}/T$ and $\bar{y}_p = \sum_{\nu=1}^{M} \bar{y}_p^{\nu}/M$. We introduce the within-class and between-class covariance matrices Σ_W and Σ_B for $x^{\nu,\tau}$ as

$$\Sigma_W = \frac{1}{MT}\sum_{\nu=1}^{M}\sum_{\tau=1}^{T}(x^{\nu,\tau} - \bar{x}^{\nu})(x^{\nu,\tau} - \bar{x}^{\nu})',$$

$$\Sigma_B = \frac{1}{M}\sum_{\nu=1}^{M}(\bar{x}^{\nu} - \bar{x})(\bar{x}^{\nu} - \bar{x})',$$

where $\bar{x}^{\nu}$ is the average of $x^{\nu,\tau}$ in the ν-th class, and $\bar{x}$ the average of whole vectors $x^{\nu,\tau}$. Then, ρ_p and σ_p^2 can be written as

$$\rho_p = (a^p)'\Sigma_W a^p, \qquad \sigma_p^2 = (a^p)'\Sigma_B a^p.$$

The lifting parameters are contained in Σ_W and Σ_B. It is preferable to make ρ_p small, and σ_p^2 large. This leads to the maximization problem

$$J(\lambda^{d,p}, \lambda^{e,p}) = (a^p)'\Sigma_B a^p - \gamma(a^p)'\Sigma_W a^p \quad \longrightarrow \quad \max., \tag{14}$$

where $\lambda^{d,p} = (\lambda_{-L}^{d,p}, ..., \lambda_L^{d,p})$, $\lambda^{e,p} = (\lambda_{-L}^{e,p}, ..., \lambda_L^{e,p})$, and γ is a penalty constant. This problem can be solved by

$$\frac{\partial J(\lambda^{d,p}, \lambda^{e,p})}{\partial \lambda_l^{d,p}} = 0, \qquad \frac{\partial J(\lambda^{d,p}, \lambda^{e,p})}{\partial \lambda_l^{e,p}} = 0, \qquad l = -L, ..., L$$

using the Newton's method. This is our learning algorithm, which is illustrated in Fig. 1.

4 Recognition System

4.1 Automatic Extraction of Training and Query Objects

This paper focuses on the recognition of vehicle types such as sedan, taxi, van, truck, and bus, moving on streets and intersections. First, images including moving objects are extracted from the difference images between adjacent video frames. The extracted images may contain shadows of vehicles, walkers, part of buildings, white lines on streets as well as vehicles themselves. So, we want to extract only vehicles by employing the proposed cosine-maximization method [6–9]. We briefly describe our object detecting process.

Proceedings of the ICVS International Cognitive Vision Workshop – ICVW 2007
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

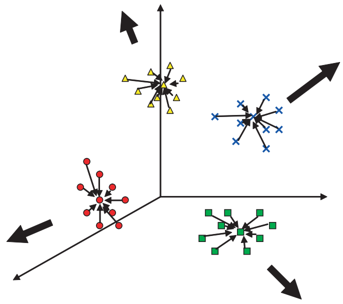


Fig. 1. Learning of lifting parameters in the feature space

1. Compute difference images from adjacent video frames.
2. Select centres of vehicles from the difference images, and learn the lifting dyadic wavelet filter at the centres, exploiting our cosine-maximization method.
3. By applying the learned filter to another set of difference images, detect the points for which cosine-values are large.
4. Extract objects around the points from the difference images by the dilation operation.
5. Map each of the extracted object regions into the corresponding original video image, and cut off a rectangular region containing it.
6. Put the values of pixels 0 in the rectangular region except for the extracted object.

4.2 Leaning Process

We automatically extract M classes of vehicle images using the extraction process described in Section 4.1. Each class includes T vehicle images of the same type. From these training images, autocorrelation vectors are constructed following the method presented in Section 2.2. The autocorrelation vectors are used for determining the lifting parameters by the learning algorithm detailed in Section 3.

4.3 Recognition Process

The recognition of vehicle types is performed as follows: A vehicle is extracted from video frames exploiting the extracting method provided in Section 4.1. From this vehicle image, we construct M_1 autocorrelation vectors $x^{qer,p}$ $p = 1, ..., M_1$ using the lifting parameters learned for the training images. Furthermore, we compute a feature vector

$$y^{qer} = (a^1 \cdot x^{qer,1}, a^2 \cdot x^{qer,2}, ..., a^{M_1} \cdot x^{qer,M_1})',$$

where a^p, $p = 1, ..., M_1$ are the discriminant vectors obtained for the training set. The type of the query vehicle is recognized by measuring the cosine distance between y^{qer} and the feature vectors for the training images.

5 Experiments

We capture many images containing vehicles moving on the street in our campus, by using the web camera equipped at the window of our laboratory. From these images, vehicles, each of which has 64×64 size, are extracted by the method described in Section 4.1. The extracting process is shown in Fig. 2. Figure 3

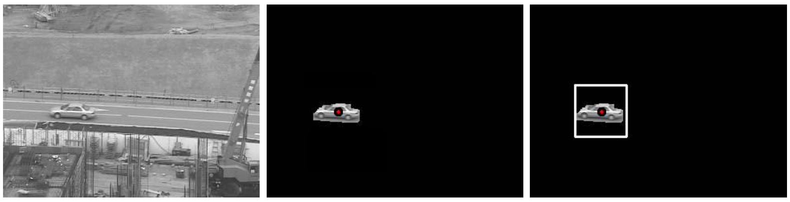


Fig. 2. Extracting process.

shows several extracted vehicles.

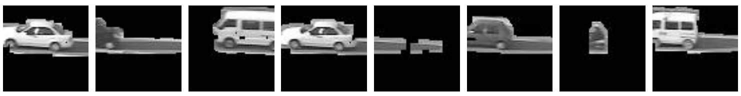


Fig. 3. Examples of extracted vehicles.

A part of the extracted vehicles is used as a training set, which is illustrated in Fig. 4. The types of vehicles are minicar, sedan, van, truck and bus, but we distinguished facing left vehicles from facing right ones, except for bus. Therefore, we have 9 classes of vehicles.

We compute the autocorrelation vectors of these training images, from which their discriminant vectors are computed by the discriminant analysis. Each of the training images is decomposed using the spline dyadic wavelet filters to get low-pass and high-pass images. From these images, lifted high-pass images are generated and the lifting parameters therein are learned following the learning algorithm detailed in Section 3.2. The number of lifting parameters in each direction is 3, i.e. $L = 1$.

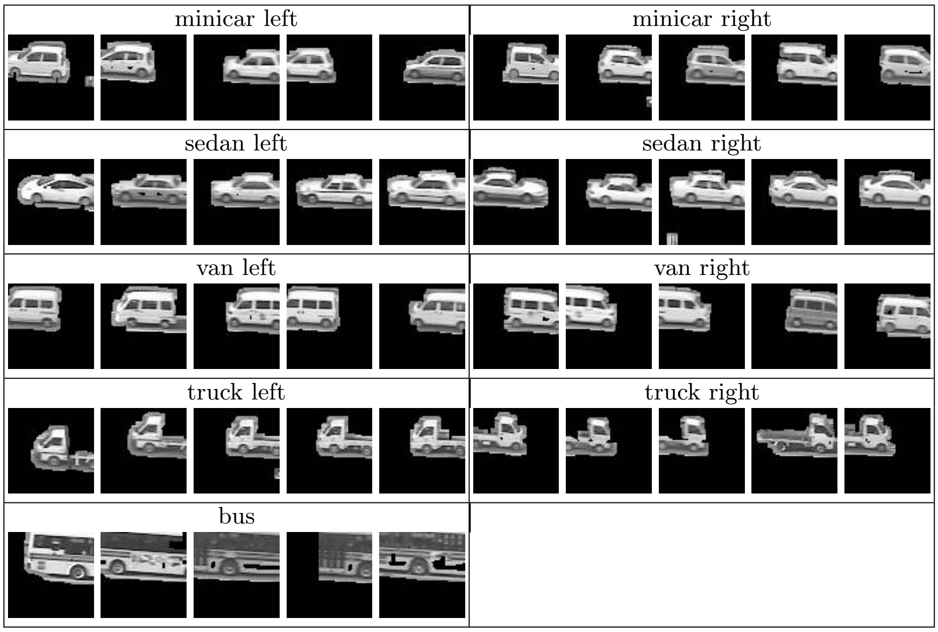


Fig. 4. Training images.

We applied the discriminant analysis to the autocorrelation vectors of the test images. We also examined our method for the test images. Figure 5 shows that our method is successful. Figure 6 is a failure case.

6 Conclusion

We have proposed a cognitive vision system based on the learning of lifting wavelets. Our experiments show that the extraction of vehicles is succeeded, but we have many misrecognition results in both the discriminant analysis and our method. There are still open problems. We need another set of image features in addition to the autocorrelation vectors. The discriminant analysis is usually used for separating objects into two classes. These are future work.

References

1. Otsu, N., Kurita, T.: A new scheme for practical, flexible and intelligent vision systems. Proc. IAPR Workshop Comp. Vis. (1988) 431-435
2. Papageorgio, C., Poggio, T.: A trainable system for object detection. Int. J. Comp. Vis. **38** (2000) 15–33

Proceedings of the ICVS International Cognitive Vision Workshop – ICVW 2007
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany

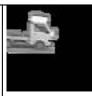
Query					
Discriminant analysis	fl van	fr van	fr minicar	fl truck	fr truck
Our method	fr truck	fr truck	fr truck	fr truck	fl minicar
Query					
Discriminant analysis	fr sedan	fl van	fl minicar	fl minicar	fr minicar
Our method	fr truck	fr truck	fr truck	fr truck	fr truck

Fig. 5. Case of success.

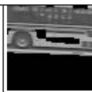
Query					
Discriminant analysis	fl van	fl minicar	fl minicar	fl bus	fl sedan
Our method	fr truck	fl sedan	fl sedan	fr truck	fl minicar

Fig. 6. Case of failure.

3. Schneiderman, H., Kanade, T.: Probabilistic modelling of local appearance and spatial relationships for object recognition. Proc. IEEE Int. Conf. Comp. Vis. Pat. Rec. (2002) 45–51
4. Sun, Z., Bebis, G., Miller, R.: Quantized wavelet features and support vector machines for on-road vehicle detection. Proc. IEEE Int. Conf. Cont. Auto. Vis. (2002) CD-ROM
5. Sun, Z., Bebis, G., Miller, R.: On-road vehicle detection: A review. IEEE Trans. Pat. Anal. Mach. Int. **28** (2006) 694–711
6. Takano, S., Niijima, K., Abdukirim, T.: Fast face detection by lifting dyadic wavelet filters. Proc. IEEE Int. Conf. Image Proc. (2003) 893–896
7. Takano, S., Niijima, K., Kuzume, K.: Personal identification by multiresolution analysis of lifting dyadic wavelets. Proc. Eur. Sig. Proc. Conf. (2004) 2283–2286
8. Takano, S., Niijima, K.: Person identification using fast face learning of lifting dyadic wavelet filters. Proc. Int. Conf. Comp. Rec. Sys. (2005) 815–823
9. Takano, S., Niijima, K.: Robust person authentication using dyadic wavelet filters learned by cosine-maximization. DOI Tech. Rep. **227** Dep. of Info. Kyushu Univ. (2006) 1–12
10. Webster, M., Welling, M., Perona, P.: Unsupervised learning of models for recognition. Eur. Conf. Comp. Vis. (2000) 18–32