Moving Object Recognition Using Wavelets and Learning of Eigenspaces

Takano, Shigeru
Department of Informatics Kyushu University

Minamoto, Teruya
Department of Informatics Kyushu University

Niijima, Koichi
Department of Informatics Kyushu University

出版情報：DOI Technical Report. 151, 1998-06-22. Department of Informatics, Kyushu University
バージョン：
権利関係：

DOI-TR-151

DOI Technical Report

Moving Object Recognition Using Wavelets and Learning of Eigenspaces

by

S. Takano, T. Minamoto, and K. Niijima

June 22, 1998

Department of Informatics
Kyushu University
Fukuoka 812-81, Japan

Email: takano@i.kyushu-u.ac.jp Phone: +81-92-583-7633

Moving Object Recognition Using Wavelets and Learning of Eigenspaces

Shigeru Takano, Teruya Minamoto, Koichi Niijima
Department of Informatics
Kyushu University
Fukuoka 816-8580
Japan
{takano, minamoto, niijima}@i.kyushu-u.ac.jp

Abstract

This paper proposes a method for recognizing moving objects, which is based on a wavelet decomposition technique and learning of eigenspaces. High frequency vectors for reference image sequences are constructed using a wavelet decomposition formula. These high frequency vectors express the characteristics of moving objects in the image sequences. From them, a covariance matrix is made and eigenvectors of this matrix are computed. Some parametric eigenspaces are learnt by these eigenvectors. Next, orthogonal projections of the reference image sequences on the parametric eigenspaces are produced. To recognize a moving object included in an unknown image sequence, a high frequency vector and its orthogonal projections on the parametric eigenspaces are constructed. The recognition is carried out by computing the distance between the projection for the unknown image sequence and that for the reference image sequences.

Keywords: Wavelets; Parametric eigenspace representation

1 Introduction

Recognition of moving objects is one of the most important problems in computer vision. This problem has many applications such as individual recognition, gesture recognition and lip reading.

So far, a number of recognition methods have been developed. The optical flow technique [3] is well known as one of such methods. This technique is used to capture the behavior of moving objects. Optical flow is calculated by the generalized gradient method based on spatio-temporal filtering, but it must be uniform inside a target image. This restriction is a bottleneck of optical flow methods. Besides this method, there are some other recognition methods which can be applied in the case that a moving object consists of some parts linked each other, for example, such as human body. However, these methods depend on the purpose.

Murase and Sakai [5] proposed a method using parametric eigenspace representation [2] to realize efficient recognition of moving objects. This method is independent of the applications and requires a little time to recognize the moving objects. However, it is needed to normalize the position and size of silhouettes extracted from moving images.

In this paper, we present a method combined a wavelet decomposition technique [4] with learning of eigenspaces to recognize moving objects. In the first, we apply a wavelet decomposition method to an image sequence in time direction, and obtain its low and high frequency components. These high frequency components express the characteristics of moving objects. Based on some high frequency vectors, we make a covariance matrix, and calculate eigenvalues and eigenvectors of this matrix. The first several eigenvectors represent the characteristics of a set of the high frequency vectors obtained. From these eigenvectors, we construct parametric eigenspaces which characterize moving objects. We can say that these eigenspaces were learnt from image sequences. To recognize a moving object from an image sequence, we make a high frequency vector from the image sequence, and construct an orthogonal projection of the vector on the parametric eigenspace. The recognition can be done by computing the distance between the orthogonal projection and those obtained from the high frequency reference vectors.

2 Wavelet decomposition

Let Z denote a set of integers, and $\{c_j^1\}_{j\in Z}$ be a sequence of real numbers. Introducing scaling coefficients α_k and wavelet coefficients β_k and using multiresolution analysis in a wavelet theory, we can obtain the following transforms:

$$c_i^0 = \sum_{k\in\mathbf{Z}} \alpha_k c_{2i+k}^1, \tag{1}$$

$$d_i = \sum_{k\in\mathbf{Z}} \beta_k c_{2i+k}^1. \tag{2}$$

The transforms (1) and (2) mean the extraction of low frequencies $\{c_i^0\}$ and high frequencies $\{d_i\}$ from the given sequence $\{c_j^1\}$, respectively. Since these transforms represent the down-sampling by a factor of two, the length of $\{c_i^0\}$ and $\{d_i\}$ is half in comparison with the original sequence $\{c_j^1\}$. The above transforms are called Mallat's transformation.

The coefficients α_k and β_k play an important role in wavelet decomposition. These coefficients are determined depending on scaling and wavelet functions. In this paper, we use Daubechies' compactly supported wavelets with tap 2. The scaling and wavelet coefficients of this wavelet are shown in Table1.

Table 1: Daubechies' scaling and wavelet coefficients

k	α_k	β_k
0	0.48296291	−0.12940952
1	0.83651630	−0.22414387
2	0.22414387	0.83651630
3	−0.12940952	−0.48296291

3 Feature extraction of moving objects

There are many approaches for extracting the feature of moving objects. One of them is a method which computes a difference image between a target image containing moving objects and its background image. To extract the feature with high quality, however, we must carry out various preprocessing for the obtained difference image such as de-noising and normalization of the size of images.

To extract the feature of moving objects, we adopt in this paper a wavelet decomposition method which is often used recently in many application fields of image analysis. Wavelet transforms are applied usually for a static picture in space direction ([1]). In this paper, we regard moving images as image sequences, and apply the wavelet decomposition in Section 2 to them for each pixel in time direction. Although both of low and high frequencies are computed for an image sequence, we adopt only high frequency components. The values of these components reflect the speed of moving objects, and are invariant with respect to the brightness of images. Our approach does not need any preprocessing for the extraction of moving objects unlike another methods.

We denote an image at time t by a vector $C^1(t)$ as

$$C^1(t) = (c_1^1(t), c_2^1(t), \cdots, c_N^1(t))^{\mathrm{T}},$$

where N is the number of pixels of the image and the symbol T indicates a transpose. For each component $c_i^1(t)$, we apply the wavelet decomposition to get

$$\begin{aligned} c_i^0(t) &= \sum_{k\in\mathbf{Z}} \alpha_k c_i^1(2t+k), \\ \tilde{d}_i(t) &= \sum_{k\in\mathbf{Z}} \beta_k c_i^1(2t+k). \end{aligned} \tag{3}$$

By the transform (3), we can extract rapid motion from the time series $\{c_i^1(t)\}_{t=0,1,\cdots}$. We see from (3) that an odd shift of the time series does not imply the shift of high frequencies $\tilde{d}_i(t)$. Thus we consider two kinds of high frequencies

$$\begin{aligned} \tilde{d}_i(t) &= \sum_{k\in\mathbf{Z}} \beta_k c_i^1(2t+k), \\ \hat{d}_i(t) &= \sum_{k\in\mathbf{Z}} \beta_k c_i^1(2t+1+k). \end{aligned}$$

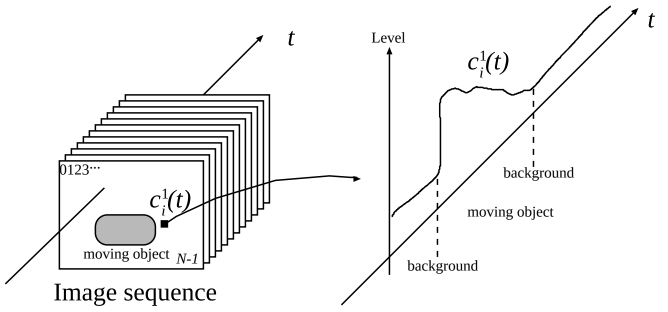


Figure 1: An image sequence in time direction

Using $\tilde{d}_i(t)$ and $\hat{d}_i(t)$, we define $d_i(t)$ as follows:

$$d_i(t) = \begin{cases} \tilde{d}_i([t/2]), & t : \text{even}, \\ \hat{d}_i([t/2]), & t : \text{odd}, \end{cases}$$

where $[\cdot]$ is the Gaussian symbol. We put

$$D(t) = (d_1(t), d_2(t), \cdots, d_N(t))^{\mathrm{T}}.$$

The vector $D(t)$ represents a feature of a moving object at time t. We call $D(t)$ a high frequency image from now on.

Figure 2: Example of $D(t)$

4 Learning of eigenspace

Using the method in Section 3, we make M high frequency reference images

$$D^m(t) = (d_1^m(t), d_2^m(t), \cdots, d_N^m(t))^{\mathrm{T}}, \qquad m = 0, 1, \cdots, M-1.$$

Based on these images, we construct an eigenspace for recognizing moving objects. To do so, we compute a covariance matrix for the reference images.

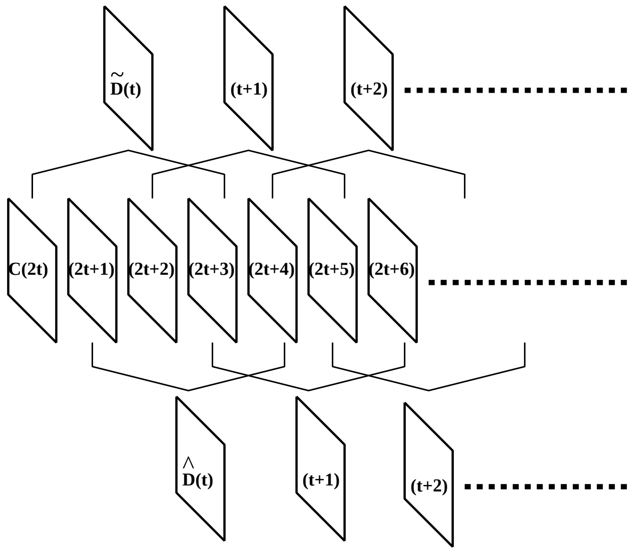


Figure 3: Wavelet decomposition in time direction

The covariance matrix for the images $D^m(t)$ is represented by

$$Q = \sum_{m=0}^{M-1}\sum_{t=0}^{T-1}(D^m(t) - \bar{D})(D^m(t) - \bar{D})^{\mathrm{T}},$$

where $\bar{D}$ is the mean vector

$$\bar{D} = \frac{1}{MT}\sum_{n=0}^{M-1}\sum_{t=0}^{T-1} D^n(t).$$

The size of matrix Q is $N \times N$.

Eigenspaces we desire can be obtained by computing eigenvectors of Q. These eigenvectors are computed using some numerical methods such as the power method and the QR decomposition method. As well known, the relation of eigenvalue and eigenvector is as follows:

$$Qe_j = \lambda_j e_j,$$

where e_j is the eigenvector for an eigenvalue λ_j. Since Q is non-negative definite, we have

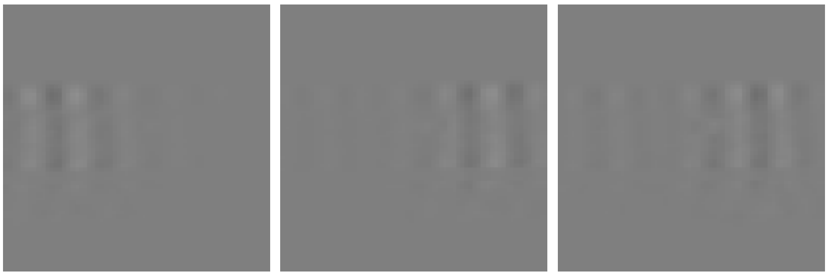
$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0.$$

A subspace spanned by k eigenvectors $\{e_1, e_2, \cdots, e_k\}$ is called a k-dimensional parametric eigenspace. Figure 4 shows first three eigenvectors.

Generally, the feature of moving objects appears in the first several eigenvectors. This means that $D^m(t)$ can be approximated by these eigenvectors.

Since $\dim(D^m(t)) = N$ and $\{e_j\}_{j=1,2,\cdots,N}$ are linearly independent, $D^m(t)$ can be expanded as

$$D^m(t) = \sum_{j=1}^{N} v_j^m(t) e_j. \tag{4}$$



e_1 e_2 e_3

Figure 4: Eigenvectors

Putting

$$V^m(t) = (v_1^m(t), v_2^m(t), \cdots, v_N^m(t))^{\mathrm{T}},$$

(4) may be written as

$$D^m(t) = (e_1, e_2, \cdots, e_N)V^m(t)$$

which implies

$$V^m(t) = (e_1, e_2, \cdots, e_N)^{\mathrm{T}} D^m(t). \tag{5}$$

The vector $V^m(t)$ characterizes $D^m(t)$ in the N-dimensional eigenspace. However, since N is usually very large, we approximate $V^m(t)$ by lower dimensional vectors

$$V^{m,k}(t) = (e_1, e_2, \cdots, e_k)^{\mathrm{T}} D^m(t). \tag{6}$$

The computation of $V^{m,k}(t)$ means the learning of $D^m(t)$ in the k-dimensional eigenspace.

5 Recognition of moving objects in eigenspace

The vector $V^{m,k}(t)$ forms a trajectory in the k-dimensional eigenspace. The M trajectories obtained by learning have already been memorized. Comparing a trajectory for an input moving image with the memorized trajectories, we can recognize the movement of the input image. This is called a method of parametric eigenspace representation.

Let $X(t)$ be a high frequency image constructed from an input image. According to (6), we define $Z^k(t)$ by

$$Z^k(t) = (e_1, e_2, \cdots, e_k)^{\mathrm{T}} X(t)$$

which represents a projection of $X(t)$ into the k-dimensional eigenspace.

To measure the difference between $Z^k(t)$ and $V^{m,k}(t)$, we define a distance of the following form

$$I^{m,k} = \min_{a,b} \sum_{t=0}^{T-1} |Z^k(t) - V^{m,k}(at+b)|,$$

where a and b are used to adjust time stretching and shifting of input images. The parameters a and b are determined depending on $Z^k(t)$.

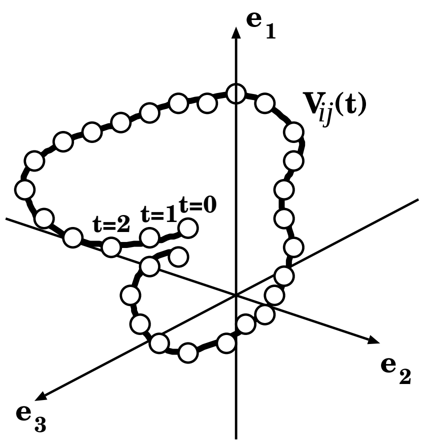


Figure 5: Parametric eigenspace, $(k = 3)$

6 Experiment

For experiment, we captured gait patterns of six people, ten times each. The

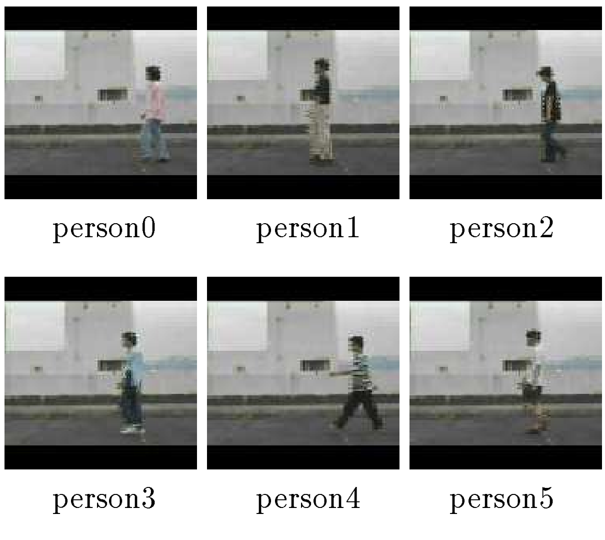


Figure 6: The gait patterns

sampling rate is 10 frames per second, and the original image size is 64×64. In our method, it is not necessary to extract moving parts from the original images, and to normalize their brightness. We used 5 gait patterns per person for learning parametric eigenspaces, and the remaining patterns for recognition test. We made three kinds of eigenspaces, concretely, with dimensions 16, 32, and 64. Recognition test was executed by using images which are not utilized for learning. Table 2 shows the recognition results. These results indicate that the larger dimensional number k is, the higher accuracy of recognition is. However, for large k, we need much time to compute the distance $I^{m,k}$, and to recognize as well. We show time for recognition

Table 2: The recognition results

	$k = 16$	$k = 32$	$k = 64$
person0	60%(3/5)	100%(5/5)	100%(5/5)
person1	40%(2/5)	80%(4/5)	100%(5/5)
person2	20%(1/5)	80%(4/5)	100%(5/5)
person3	40%(2/5)	60%(3/5)	60%(3/5)
person4	0%(0/5)	100%(5/5)	100%(5/5)
person5	60%(3/5)	40%(2/5)	40%(2/5)
all	36%(11/30)	77%(23/30)	83%(25/30)

in Table 3. Table 3 says that fast recognition can be realized in small dimensional parametric eigenspaces.

Table 3: The recognition time on the computer SGI Octane

$k = 16$	$k = 32$	$k = 64$
3.0sec	5.0sec	11.0sec

7 Conclusion

We proposed a moving object recognition method of using a wavelet decomposition and learning of eigenspaces. The characteristics of a moving object were extracted by applying the wavelet decomposition of image sequences in time direction.

In the experiment, we constructed eigenspaces based on video images of human gaits. The experiment results show that our method is effective to extract motion characteristics and to recognize moving objects.

We used in this paper Daubechies' wavelet coefficients to design eigenspaces. Since the eigenspace depends on the choice of wavelet coefficients, we have to construct it by using another type of wavelet coefficients and to compare new recognition results with those in Section 6. This is a problem to be resolved in the future.

References

[1] E. J. Stollnitz, T. D. Derose, D. H. Salesin : *Wavelets for Computer Graphics*, Morgan Kaufmann Publishers, Inc. San Francisco, California, 1996.

[2] E.Oja : *Subspace Methods for pattern Recognition*, Research Studies Press, Herfordshire, 1983.

[3] G.Adiv : Determining three-dimensional motion and structure from optical flow generated by several moving objects, IEEE Trans. Vol. PAMI-7, NO.4, pp384-401, 1985.

[4] I.Daubechies : *Ten Lectures on Wavelet,* New York: SIAM Press, 1992.

[5] H.Murase, R.Sakai : Moving object recognition in eigenspace representation: gait analysis and lip reading, Pattern Recognition Letters 17, pp155-162, 1996.