EURASIP Journal on Applied Signal Processing 2004:14, 2113–2131

RST-Resilient Video Watermarking Using Scene-Based Feature Extraction

Han-Seung Jung
School of Electrical Engineering and Computer Science, Seoul National University, San 56-1, Sillim-Dong, Gwanak-gu, Seoul 151-742, Korea
Email: jhs@ipl.snu.ac.kr

Young-Yoon Lee
School of Electrical Engineering and Computer Science, Seoul National University, San 56-1, Sillim-Dong, Gwanak-gu, Seoul 151-742, Korea
Email: yylee@ipl.snu.ac.kr

Sang Uk Lee
School of Electrical Engineering and Computer Science, Seoul National University, San 56-1, Sillim-Dong, Gwanak-gu, Seoul 151-742, Korea
Email: sanguk@ipl.snu.ac.kr

Received 31 March 2003; Revised 5 April 2004

Watermarking for video sequences should consider additional attacks, such as frame averaging, frame-rate change, frame shuffling or collusion attacks, as well as those of still images. Also, since video is a sequence of analogous images, video watermarking is subject to interframe collusion. In order to cope with these attacks, we propose a scene-based temporal watermarking algorithm. In each scene, segmented by scene-change detection schemes, a watermark is embedded temporally to one-dimensional projection vectors of the log-polar map, which is generated from the DFT of a two-dimensional feature matrix. Here, each column vector of the feature matrix represents each frame and consists of radial projections of the DFT of the frame. Inverse mapping from the one-dimensional watermarked vector to the feature matrix has a unique optimal solution, which can be derived by a constrained least-square approach. Through intensive computer simulations, it is shown that the proposed scheme provides robustness against transcoding, including frame-rate change, frame averaging, as well as interframe collusion attacks.

Keywords and phrases: scene-based video watermarking, RST-resilient, radial projections of the DFT, feature extraction, inverse feature extraction, least-square optimization problem.

1. INTRODUCTION

The widespread utilization of digital data leads to illegal use of copyrighted material, that is, unlimited duplication and dissemination via the Internet. As a result, this unrestricted piracy makes service providers hesitate to offer services in digital form, in spite of the digital audio and video equipment replacing the analog ones. In order to overcome this reluctancy and possible copyright issues, the intellectual property rights of digitally recorded material should be protected. For the past few years, the copyright protection problems for digital multimedia data have drawn a significant interest with the increased utilization of the Internet.

In order to protect the copyrighted multimedia data, many approaches, including authentication, encryption, and digital watermarking, have been proposed. The encryption methods may guarantee secure transmission to authenticated users via the defective Internet. Once decrypted data, however, is identical to the original and its piracy cannot be restricted. The digital watermarking is an alternative to deal with these unlawful acts. Watermarking approaches hide invisible mark or copyright information in digital content and claim the copyright. The mark should be robust enough to survive legal or illegal attacks. It is also desirable that some illegal attempts should suffer from the degradation in visual quality, without erasing the watermarks.

For an effective watermarking scheme, two basic requirements should be satisfied: transparency and robustness. Transparency means the invisibility of watermarks embedded in image data without degrading the perceptual quality by watermarking. Robustness means that the watermark should not be removed or detected by attacks, that is, signal

processing, compression, resampling, cropping, geometric distortion, and so forth. Many watermarking algorithms for images have been developed, which are generally categorized into spatial-domain [1, 2, 3] and frequency domain techniques [4, 5, 6, 7, 8, 9, 10]. In most cases, image watermarking techniques in frequency domain, such as discrete cosine transform (DCT), discrete Fourier transform (DFT), and wavelet transform, are preferred because of their efficiency in both robustness and transparency. Specifically, from the viewpoint of geometric attacks, DFT-based or template-embedding watermarking algorithms yield better performance than the others in general [7, 8, 9].

In case of video watermarking, new kinds of attacks are available to remove the marks. These attacks include frame-averaging, frame-rate change, frame swapping, frame shuffling, and interframe collusion. Since video signals are highly correlated between frames, the mark in video is vulnerable to these attacks, which affect the mark adversely without degrading video quality severely. Since frames in a scene are so analogous, completely different watermarks in each frame may be detected and removed easily by a simple collusion scheme. Also, in case of applying an identical watermark to the whole video sequence, this mark can be easily estimated without satisfying the statistical invisibility. So, many video watermarking algorithms address these collusion issues [11, 12, 13]. Video sequences are composed of consecutive still images, which can be independently processed by various image watermarking algorithms. In this case, interframe collusions should be considered as in [11]. Also, three-dimensional (3D) transforms are good approaches for the video watermarking since they can be easily generalized from two-dimensional (2D) techniques for images and are robust against collusion attacks [12, 13, 14]. Watermarking in bit-stream structure can be another solution for video watermarking [15, 16, 17, 18], but this approach may be vulnerable to re-encoding or transcoding.

In this paper, we present a novel video watermarking algorithm of feature-based temporal mark embedding strategy. Video sequence consists of a number of scene segments, and each scene may be a good temporal watermarking unit because the scene itself is always available after attacks of frame-rate change, frame swapping, frame shuffling, and so forth. In many cases, illegal distributors transcode the original video sequence to others, for example, re-encoding MPEG-2 video to MPEG-4, and generally this process forces the original data to suffer from the aforementioned attacks. Thus, we employ features extracted from each video scene as a watermarking domain. The watermark embedding procedure is composed of three steps: feature extraction, watermarking in feature domain, and inverse feature extraction. First, scene-change detection algorithms divide a video sequence into scenes using luminance projection vectors (LFVs) [19, 20]. In each scene, one-dimensional (1D) frequency projection vectors (FPVs), which represent the characteristics of the frames, are extracted. An FPV is obtained from the radial-wise sum of log-polar map, generated from the DFT of the frame. This 1D FPV is known to be invariant to rotation, scaling, and translation (RST) [7, 8, 9]. Then all these vectors in a scene compose a 2D matrix, which is interpolated on the temporal axis and becomes projection vector time flow matrix (PVTM). Specifically, for an $N \times M$ PVTM, M is the length of predefined time flow and N is the length of the FPV. Secondly, a watermark is embedded in a 1D watermarking feature vector (WFV), which is generated from the PVTM using the same process of obtaining the FPV. In the proposed algorithm, scalings in image and temporal domains mean aspect-ratio change of frames and frame-rate change, respectively. Thus, this temporal mark embedding strategy is expected to be invariant to some video-oriented attacks. Moreover, since the embedding approach is not a one-to-one mapping, inverse feature extraction should be considered. We find that constrained linear least-square method can achieve the global minimum of the optimization problem, and the inverse mapping from the watermarked feature vector (WFV) to the PVTM has a unique optimal solution.

This paper is organized as follows. In Section 2, we present the efficiency and reasonability of temporal watermarking for video sequences. Then, the proposed algorithm is described in Section 3, where we present the watermark embedding and detection procedures. In Section 4, the inverse feature extraction, which inversely maps the watermark in the feature domain to the original video frame domain, is derived. Section 5 examines the performance of the proposed algorithm, and shows that the proposed scheme yields a satisfying performance, both in terms of transparency and robustness. In Section 6, we present the conclusion of this paper.

2. TEMPORAL WATERMARKING FOR VIDEO SEQUENCE

Since a typical video sequence is composed of many frames with temporal redundancy, statistical watermark-estimation, collecting and analyzing the video frames can be an effective attack against video watermarking. Frames within a scene are highly correlated. So, one can exploit the temporal redundancy, either of the frames or of the watermark, to estimate and remove the watermark signal. This collusion has become an important issue for video watermarking. Su et al. [11] have defined two types of linear collusion attacks. One is due to a fixed watermark pattern in large numbers of visually distinctive video frames, and the other is due to independent watermark patterns in large numbers of visually similar frames. Based on the statistical analysis of linear collusions, they presented a spatially localized image-dependent framework for collusion-resilient video watermarking. Frame-based watermarking algorithms are employed, and it is shown that the spatial domain approach outperforms the DFT approach in case of severe compression, while the DFT approach is more robust to general attacks.

Alternative approaches, based on the idea of temporal watermarking, are available to cope with the interframe collusion and consider frames in a sequence jointly. Most of these algorithms are generally based on the extended versions of 2D transforms, that is, 3D DFT or 3D wavelet transform.

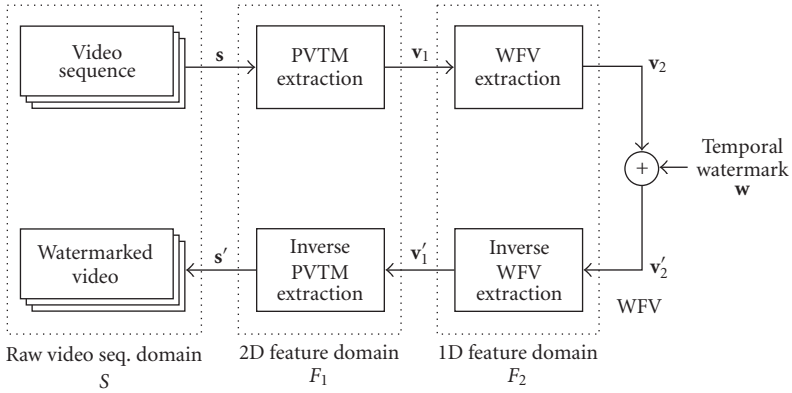


Figure 1: Framework of the proposed algorithm.

Swanson et al. [12] proposed a scene-based video watermarking algorithm using a temporal wavelet transform of the video scenes. A wavelet transform along the temporal domain of a video scene results in a multiresolution temporal representation of the scene: static (lowpass) and dynamic (highpass) video components. They also used perceptual models for an invisible and robust watermark. Deguillaume et al. [13] employed the 3D DFT in which a watermark and a template are encoded in the 3D DFT magnitude of video sequence and in the log-log-log map of the 3D DFT magnitude, respectively. These algorithms are also resilient to the temporal modifications of frame-rate change, frame swapping, frame dropping, as well as frame-based degradation and distortion.

Temporal watermarking strategy must be reliable against such attacks. A scene can be a good segment unit in temporal domain as in [12]. Scenes are always maintained in spite of the aforementioned temporal attacks. So, the proposed algorithm is based on the idea of temporal mark embedding, but it is not just an extension of 2D transforms. The proposed algorithm uses a new feature domain for watermarking. The feature-based watermarking facilitates the 3D problem of temporal mark-embedding and real-time mark detection while providing the resilience against collusion and temporal attacks.

3. PROPOSED ALGORITHM

3.1. Feature space for video watermarking

In many watermarking systems, the watermarks are embedded in the transform domain, such as DFT domain or DCT domain. That is, these systems use the transform domain as the watermark space, in which the watermarks are inserted and detected [4]. In these cases, the dimension of the watermark domain is the same as that of the media space. For video watermarking, simple extensions of these transforms have been applied to video sequences in [12, 13, 14]. These algorithms provide effective performance against interframe collusion and noise-like attacks. In this paper, however, we employ the feature domain as the watermark space. The feature has two meanings: some summarization of video contents and a 1D mark embedding vector derived from the watermark signals. We modify the feature according to the watermark signals. The proposed algorithm has the following three advantages.

(1) Complexity: the dimension of a video sequence is often too large, so we use the feature as the watermark domain, which has a reduced dimension.
(2) Robustness: the feature is RST-invariant.
(3) Transparency: we select the masking method[1] minimizing the error to achieve a good invisibility.

We have defined two types of the feature spaces; one represents frame and video contents, and the other is the watermarking space. As shown in Figures 1 and 2, the PVTM and the FPVs represent video contents in a scene and corresponding frames, respectively, and the WFV is considered as a watermarking space. Here, the FPV and the WFV have a similar structure that is RST-invariant. In [7], Lin et al. proposed an RST-resilient algorithm for the image watermarking. They defined a 1D projection of the magnitude of the Fourier spectrum, denoted by $g(\theta)$ and given by

$$g(\theta) = \sum_{j} \log\left(\left|I(\rho_j, \theta)\right|\right), \tag{1}$$

where $I(\rho_j, \theta)$ is the Fourier transform of an image $i(x, y)$ in log-polar coordinates. $g(\theta)$ is invariant to both translation and scaling, and rotations result in a circular shift of the values of $g(\theta)$. This strategy is employed basically in embedding a watermark vector to the WFV space in the proposed algorithm, except for the inverse mark embedding to the original signals. We consider this inverse problem a linear constrained problem, which will be discussed in Section 4.

In the proposed algorithm, the meanings of the RST are somewhat different from those in image watermarking algorithms. The PVTM is invariant to temporal attacks, such as frame-rate change and frame scaling, which may occur during the process of transcoding, due to the interpolation

[1] In this paper, it is called the inverse feature extraction procedure.

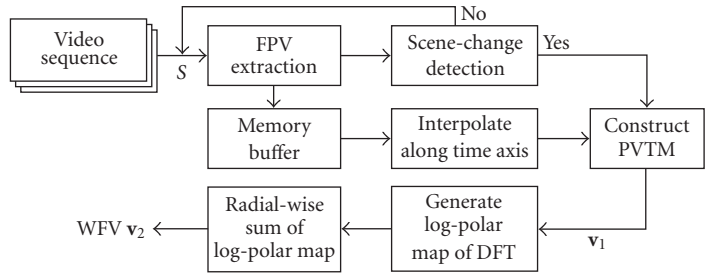


Figure 2: Construction of the WFV $\mathbf{v}_2$.

along the temporal axis in the process of constructing PVTM. The rotation in a frame yields a circular shift in the PVTM domain, which does not change the DFT magnitude of the PVTM, but changes the phase component only. The DFT magnitude itself is invariant to the translation of a frame, and moreover, the PVTM domain and its DFT magnitude are immune against the translation. For interframe collusion, the effect of PVTM is the same as the frame-rate change as mentioned before. Thus, this feature-based watermarking strategy is RST-invariant and reasonable for video watermarking, and we can expect that the proposed approach would provide the robustness against the aforementioned attacks as well as interframe collusions.

3.2. Watermark embedding

In the proposed scheme, a single-bit watermark vector of length N is embedded and detected, in which the presence of the watermark claims the ownership for the copyright material. As in Figures 1 and 2, the watermark embedding algorithm can be summarized as follows.

(1) Divide full video sequences into scenes using the distance function in [19, 20] in which the measuring functions employing the LPVs, instead of full frames, are used for efficiency. The LPV is the projection of luminance image on column or row axis. Let f_i denote an ith image of size $M \times N$ in a scene, and then the luminance projections for the nth row and the mth column, denoted by l_n^r and l_m^c, respectively, are $l_i^r(n) = \sum_{m=1}^{M} \mathrm{Lum}\{f_i(m,n)\}$ and $l_i^c(m) = \sum_{n=1}^{N} \mathrm{Lum}\{f_i(m,n)\}$. So, the dissimilarity between ith and jth frames can be defined as follows:

$$d(i,j) = \frac{1}{255 \cdot (M+N)} \left\{ \frac{1}{N} \sum_{n=1}^{N} \left| l_i^r(n) - l_j^r(n) \right| + \frac{1}{M} \sum_{m=1}^{M} \left| l_i^c(m) - l_j^c(m) \right| \right\}. \tag{2}$$

In many cases, the LPV is extracted from the DC image, which is 1/64 of the original image size [19]. This strategy can decrease the calculation complexity and also guarantee robustness against video coding.

(2) In each scene, extract the FPVs from the frames. First, each frame is put to an $l \times l$ square image, padded with trailing zeros, where l is generally confined to powers of two for the fast Fourier transform (FFT). Second, we transform the zero-padded image of the kth frame $i_k(x,y)$ into its Fourier transform $I_k(\xi_1, \xi_2)$. Next, zero-frequency component of $I_k(\xi_1, \xi_2)$ is shifted to the center of spectrum by swapping the first and third quadrants and the second and forth quadrants. Finally, the FPV of the kth frame can be obtained through applying a projection operator $\mathcal{R}$ to $|I_k(\xi_1, \xi_2)|$ given by

$$\mathbf{f}_k = \mathcal{R}\left| I_k(\xi_1, \xi_2) \right| = \{\mathbf{f}_k(i)\}, \tag{3a}$$

where

$$\mathbf{f}_k(i) = \mathcal{R}_{\theta_i}\left| I_k(\xi_1, \xi_2) \right|. \tag{3b}$$

The symbol $\mathcal{R}$, denoting the *Radon transform operator*, is also called the projection operator. For matrices, $\mathcal{R}_{\theta_i}$ is the projection operator along a radial line oriented at an angle θ_i at a specific distance from the origin. More specifically, $\mathbf{X}$ can be projected to $\mathbf{x}(i)$ for the angle θ_i, and the resulting $\mathbf{x}$ is a column vector containing the Radon operation for some prespecified degrees written as

$$\mathbf{x} \overset{\text{def}}{=} \mathcal{R}\mathbf{X} = \{\mathcal{R}_{\theta_k}\}\mathbf{X} = \{\mathbf{x}(i)\}, \tag{4a}$$

where

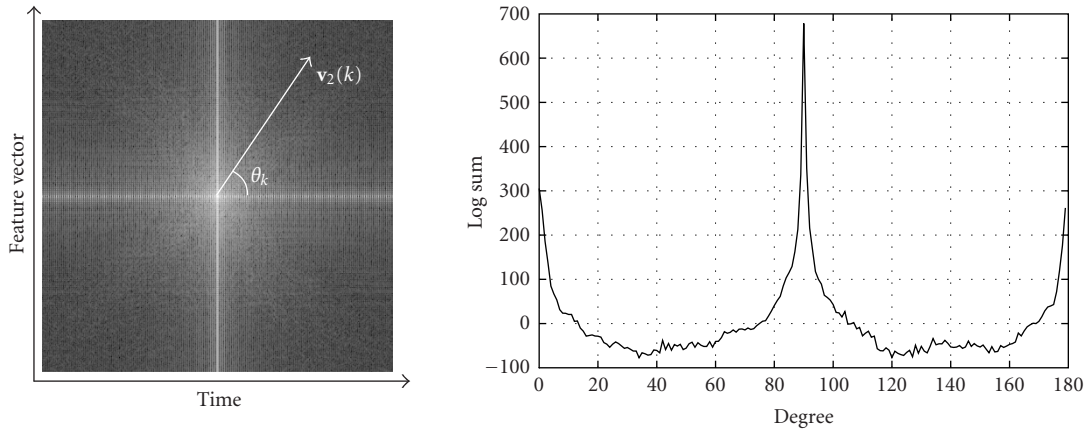
$$\mathbf{x}(i) \overset{\text{def}}{=} \mathcal{R}_{\theta_i}\mathbf{X} = \sum_{\xi_1=1}^{l} \sum_{\xi_2=1}^{l} [\mathbf{X}]_{\xi_1,\xi_2} \delta\left((\xi_1 - l')\cos\theta_i + (\xi_2 - l')\sin\theta_i\right). \tag{4b}$$

The Radon operation needs the resampling and interpolation due to the coordinate conversions, and, in this work, we adopt the bilinear interpolation.

(3) The PVTM is constructed with the group of the FPVs in a scene, more specifically, which goes through the interpolation along the temporal axis. As shown in Figure 2, the same process of step (2) is also applied to the 2D matrix PVTM, denoted by $\mathbf{V}_1$. That is, the WFV $\mathbf{v}_2$ is obtained by applying (4) to $\log|\mathbf{V}_1(\xi_1, \xi_2)|$, where $\mathbf{V}_2(\xi_1, \xi_2)$ is the DFT of $\mathbf{V}_1$. The WFV $\mathbf{v}_2$ can be written as

$$\mathbf{v}_2 = \mathcal{R}\log\left|\mathbf{V}_2(\xi_1, \xi_2)\right|, \tag{5}$$

where $\mathbf{v}_2$ is a 1D vector and we modify the vector with a watermark message by a mixing function $f_{\mathrm{wm}}(\mathbf{v}_2, \mathbf{w}_2)$.



(a) (b)

FIGURE 3: (a) The DFT magnitude of the PVTM, or an equalized image of $\log|\mathbb{V}_1|$. (b) An example of WFV with length $N = 180$.

(4) Compute the watermarked version $\mathbf{v}_2'$ using a watermark mixing function $f_{\text{wm}}(\mathbf{v}_2, \mathbf{w}_2)$ given by

$$\mathbf{v}_2' = f_{\text{wm}}(\mathbf{v}_2, \mathbf{w}_2) = \mathbf{v}_2 + \alpha \mathbf{w}_2, \tag{6}$$

where α and $\mathbf{w}_2$ are a weighting factor and the watermark message, respectively.

(5) The generated signal is in the 1D vector form, and its inverse function, that is, from a lower-dimensional space to the original Fourier magnitude, cannot be defined definitely. Also, mapping the PVTM to original video frames has a similar problem. It is often the case that linear programming can be employed in order to find the solution for these constrained problems. So, we adopt a linear programming method which will be explained in Section 4.

3.3. Watermark detection

In order to determine the presence of the watermark, in many cases, a correlation-based detection approach can be used. That is, a correlation coefficient, derived from a given watermark pattern and a signal with/without the watermark, is used to check the presence of the watermark. The watermark is determined to be present if the correlation value is larger than a specific threshold T and vice versa. This strategy is simple and effective for single-bit watermarking systems [5, 7, 21], which holds true for the proposed algorithm. Moreover, in this paper, the 1D feature vector $\mathbf{v}_2$ is adopted as a watermark space, which alleviates the complexity of the detection procedure, and thus makes the real-time detection possible.

The procedure of the watermark detection follows that of the watermark embedding. First, video segments $\mathbf{s}$ are extracted from a suspected video content $\mathbf{c}$. Then, the WFV $\mathbf{v}_2$, generated from the steps (1)–(3) of the watermark embedding procedures, is correlated with the expected watermark signals $\mathbf{w}$ to obtain the distance metric $d(\mathbf{v}_2, \mathbf{w}_2)$ given by

$$d(\mathbf{v}_2, \mathbf{w}_2) = \frac{E[\mathbf{v}_2^T \mathbf{w}_2]}{\sqrt{E[\mathbf{v}_2^T \mathbf{v}_2] E[\mathbf{w}_2^T \mathbf{w}_2]}}. \tag{7}$$

If the metric $d(\mathbf{v}_2, \mathbf{w}_2)$ is greater than a threshold T, which may be signal-dependent, the signal is declared to contain the watermark. Otherwise, the signal is declared to be not a watermarked one.

As shown in Figure 3, however, the feature vector does not satisfy the properties of the random sequence completely. So, we cannot expect that (7) yields optimum results. According to the detection theory, the correlation detectors are optimum only for a signal modeled as additive white Gaussian noise (AWGN) [4, 21, 22]. Therefore, the detection performance can be improved by making a nonwhite signal to a signal with a constant power spectrum. This can be achieved by a regression method using least-squares fitting [23]. In the proposed algorithm, the feature vector $\mathbf{v}_2$ is predicted by a kth-degree polynomial written as

$$\mathbf{v}_2 = a_0 + a_1 \mathbf{x} + \cdots + a_k \mathbf{x}^k; \tag{8}$$

and the detector uses the regression residuals $\mathbf{e}_\mathbf{v}$ of the feature vector $\mathbf{v}_2$ given by

$$\mathbf{e}_\mathbf{v} = \mathbf{v}_2 - \mathbf{X}\mathbf{a}, \tag{9}$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & \cdots & x_2^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n_{v2}} & x_{n_{v2}}^2 & \cdots & x_{n_{v2}}^k \end{bmatrix}, \tag{10}$$

$$\mathbf{a} = (\mathbf{X}^\mathbf{T}\mathbf{X})^{-1}\mathbf{X}^\mathbf{T}\mathbf{v}_2.$$

The computer simulation shows that the detection performance can be improved by the regression method.

4. INVERSE FEATURE EXTRACTION

As shown in Figure 1, the watermarking procedure is divided into two stages: generation and modification of the 1D WFV

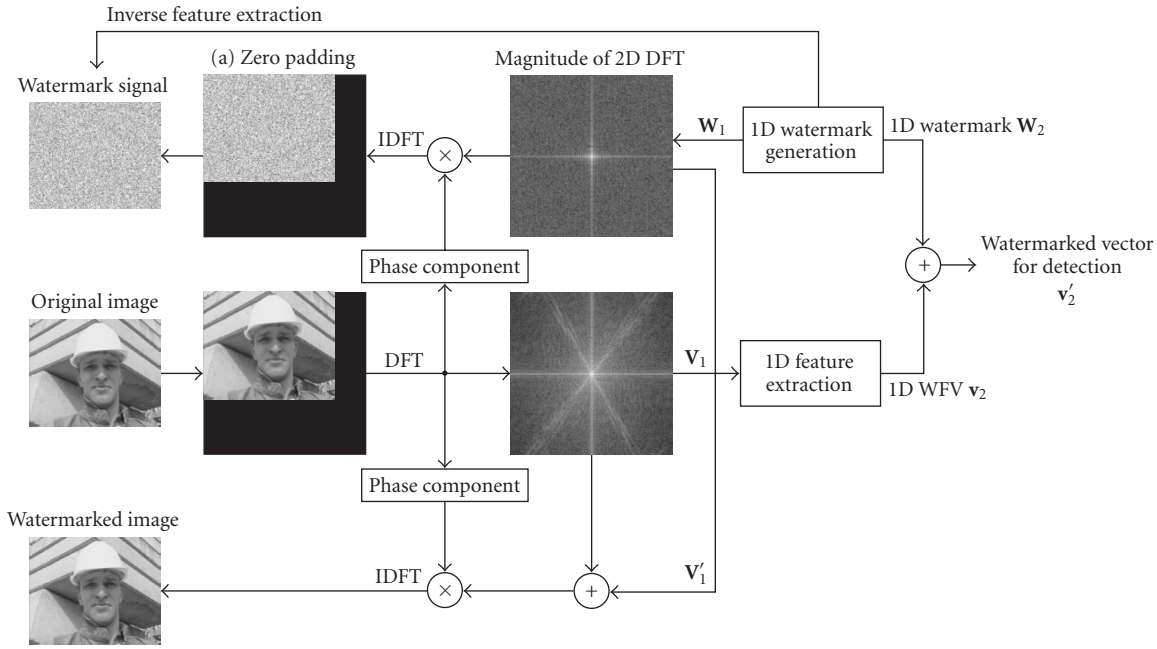


Figure 4: Inverse feature extraction.

and its inverse. The forward processing, in which the watermark vector is weighted and added to the WFV domain, is simple. However, its inverse mapping, which cannot be obtained by straightforward methods, has no unique solution. So, we approach the inverse solution using a linear programming approach. A linear programming problem is defined, as its name implies, by linear functions of the unknowns; the objective is linear in the unknowns, and the constraints are linear equalities or linear inequalities in the unknowns. In this paper, a method for the constrained linear least-square problems is adopted to find the watermark mask.

The watermarked signals $\mathbf{v}_2'$, $\mathbf{v}_1'$, and $\mathbf{s}'$ in Figure 1 are obtained in the reverse order of the feature extraction. During processing, the 1D watermark vector $\mathbf{w}_2$ is weighted and added to the WFV domain, yielding the WFV $\mathbf{v}_2'$. It is difficult to map the 1D watermarked vector $\mathbf{v}_2'$ to the 2D signal $\mathbf{v}_1'$. In the same way, it is also difficult to map $\mathbf{s}'$ to the corresponding video frame. In each domain, that is, S, F_1, and F_2, the modified signals can be represented as weighted sums of the original signal and its watermarking mask. That is, since the feature extraction and its inverse mapping are linear operations, the watermark, embedded in the WFV and mapped to video frames, can be represented in a masking form:

$$\begin{aligned} \mathbf{v}_2' &= \mathbf{v}_2 + \alpha \cdot \mathbf{w}_2 \\ \Longleftrightarrow\ \mathbf{v}_1' &= \mathbf{v}_1 + \alpha \cdot \mathbf{w}_1 \\ \Longleftrightarrow\ \mathbf{s}' &= \mathbf{s} + \alpha \cdot \mathbf{w}_0. \end{aligned} \tag{11}$$

In the inverse feature extraction, the watermark signals $\mathbf{w}_1$ and $\mathbf{w}_0$ are constructed from the watermark $\mathbf{w}_2$ in the WFV, which are in the 2D feature domain of the PVTM and the original video, respectively. So, we concentrate on only the watermark mask and not on the watermarked signals.

4.1. Inverse log-polar projection

In order to find the optimal solution of $\mathbf{w}_1$ from $\mathbf{w}_2$, we follow the forward processing. Note that the watermark $\mathbf{W}_2$ modifies only the magnitude of the Fourier transform of the cover data $\mathbf{V}_1$, as shown in Figure 4, and hence the Fourier transforms of $\mathbf{W}_1$ and $\mathbf{V}_1$ have the same phase in common. $\mathbf{W}_1$ and $\mathbf{V}_1$ can be written as

$$\begin{aligned} \mathbf{V}_1 &= \{v_1(i,j) : i = 1,\ldots,n_f,\ j = 1,\ldots,n_t\} = \mathcal{O}_c^{-1}\mathbf{v}_1, \\ \mathbf{W}_1 &= \{w_1(i,j) : i = 1,\ldots,n_f,\ j = 1,\ldots,n_t\} = \mathcal{O}_c^{-1}\mathbf{w}_1, \end{aligned} \tag{12}$$

where n_f and n_t are the size of FPV and the number of video frames in the scene, respectively, and $\mathcal{O}_c$ is a *column stacking operator* [24] given by

$$\mathbf{x} = \sum_{k=1}^{n} \mathbf{N}_k \mathbf{X} \mathbf{d}_k \stackrel{\text{def}}{=} \mathcal{O}_c \mathbf{X}, \tag{13a}$$

where

$$\begin{aligned} \mathbf{N}_k &= \{\mathbf{C}_j : \mathbf{C}_j = \delta(j-k)\mathbf{I}_m,\ j = 1,\ldots,n\}, \\ \mathbf{d}_k &= \{d_j : d_j = \delta(j-k),\ j = 1,\ldots,n\}. \end{aligned} \tag{13b}$$

A column stacking operation on an $m \times n$ matrix $\mathbf{X}$ generates a 1D $mn \times 1$ column vector $\mathbf{x}$; the (i,j) element of $\mathbf{X}$ is mapped to $(m(j-1)+i,1)$ element of $\mathbf{x}$. Each matrix is reconstructed to $l \times l$ square matrix by $\mathbf{W}_1^p = \mathbf{I}_{l,m}\mathbf{W}_1\mathbf{I}_{n,l}$ and $\mathbf{V}_1^p = \mathbf{I}_{l,m}\mathbf{V}_1\mathbf{I}_{n,l}$.

As shown in Figure 4, assuming that the PVTM is an image, the watermarked data $\mathbf{V}_1'$ can be written as $\mathbf{V}_1' = \mathbf{V}_1 + \alpha\mathbf{W}_1$ from (11). The cover data $\mathbf{V}_1$ and the unknown watermark mask $\mathbf{W}_1$ have the same dimension. The 1D watermarked vector $\mathbf{v}_2'$ for detection cannot be exactly identical with the watermarked vector $\mathbf{v}_2'$ obtained by feature extraction from the 2D matrix $\mathbf{V}_1'$. The reason is that the inverse feature extraction function from $\mathbf{w}_1$ to $\mathbf{W}_1$ is ill-conditioned and it is not practical to perform this inversion precisely. Instead, we use a linear least-square optimization method. We construct the 2D DFT magnitude $\mathbf{W}_1$ from the 1D vector $\mathbf{w}_2$ with two constraints; one is the feature extraction condition from $\mathbf{W}_1$ to $\mathbf{w}_1$, and the other is that the inverse DFT (IDFT) values of the generated $\mathbf{W}_1$, which have the same phase components as $\mathbf{V}_1$, should be zeros in zero-padding area as in Figure 4(a).

The log-polar projection of the Fourier transform of $\mathbf{W}_1$, or $\mathbb{W}_1$, should be the watermark vector $\mathbf{w}_2$, which can be written as $\mathcal{R}\log|\mathbb{W}_1| = \mathbf{w}_2$. As mentioned above, $\mathbb{W}_1$ has the same phase as $\mathbb{V}_1$ given by

$$\begin{aligned}\mathbf{W}_1(i,j) &\overset{\text{FFT}}{\rightleftharpoons} \mathcal{F}(\mathbf{W}_1(i,j)) \\ &\overset{\text{def}}{=} \mathbb{W}_1(\xi_1,\xi_2) = |\mathbb{W}_1(\xi_1,\xi_2)|\exp(j\angle\mathbb{V}_1(\xi_1,\xi_2)),\end{aligned} \tag{14a}$$

where

$$\angle\mathbb{V}_1(\xi_1,\xi_2) = \arctan\left(\frac{\Im\mathbb{V}_1(\xi_1,\xi_2)}{\Re\mathbb{V}_1(\xi_1,\xi_2)}\right). \tag{14b}$$

Thus, we define the constrained problem as

$$\mathbf{w}_1 = \arg\min\{\mathbf{w}_1^T\mathbf{H}\mathbf{w}_1 : \mathcal{R}\log|\mathbb{W}_1| = \mathbf{w}_2\}, \tag{15}$$

where $\mathbf{H}$ is a weighting factor and positive semidefinite, considering the human visual system (HVS) and the conversion from the feature domain to the DFT domain [25]. In case that the matrix $\mathbf{H}$ is an identity, the object function $\mathbf{w}_1^T\mathbf{H}\mathbf{w}_1$ becomes the Euclidean or l_2-norm of $\mathbf{w}_1$. The magnitude of low frequencies can be much larger than the magnitude of mid and high frequencies. In such case, low frequencies can be too dominant. To avoid this problem, Lin et al. sum the logs of the magnitudes of the frequencies along the columns of the log-polar Fourier transform, rather than summing the magnitudes themselves. A beneficial side effect of this is that a desired change in a given frequency is expressed as a fraction of the frequency's current magnitude rather than as an absolute value. In the proposed approach, a weighting matrix $\mathbf{H}$ can be substituted instead of the logarithm operation. This is better from a fidelity perspective.

Note that the zero padding is applied before the Fourier transform to increase the resolution. In order to obtain an optimal watermark mask, additional constraints are required besides the aforementioned one. That is, for the inverse Fourier transform of generated watermark mask with the same phase as the PVTM, the corresponding values to the region outside of the PVTM should be zeros. This strategy minimizes the loss of the energy which leaks from the image outside during IDFT. So, (15) has another constraint given by

$$\mathbf{W}_1 = \mathcal{F}^{-1}(\mathcal{S}^{-1}\mathbb{W}_1), \tag{16}$$

$$\mathbf{W}_1(i,j) = 0, \quad \text{if } i > n_f \text{ or } j > n_t. \tag{17}$$

Equation (17) can be rewritten as

$$\mathbf{W}_1 - \mathbf{T}_{n_f}\mathbf{W}_1\mathbf{T}_{n_t} = \mathbf{O}, \quad \mathbf{T}_n = \mathbf{I}_{l,n}\mathbf{I}_n\mathbf{I}_{n,l}. \tag{18}$$

Finally, from (15), (16), and (18), we have

$$\begin{aligned}\mathbf{w}_1 = \arg\min\{&\mathbf{w}_1^T\mathbf{H}\mathbf{w}_1 : \mathcal{R}\log|\mathbb{W}_1| = \mathbf{w}_2, \\ &\mathbf{W}_1 - \mathbf{T}_{n_f}\mathbf{W}_1\mathbf{T}_{n_t} = \mathbf{O}\},\end{aligned} \tag{19}$$

which is a least-square optimization problem with linear constraint equation. So, we can solve this problem using the quadratic programming [26, 27]. The construction of $\mathbf{W}_0$ from $\mathbf{W}_1$ follows the similar procedure.

4.2. Uniqueness and existence of the solution

In the proposed scheme, the feature extraction and its inverse can be formulated as a linear constrained problem given in the form

$$\min\{\mathbf{x}^T\mathbf{H}\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\}, \tag{20}$$

where $\mathbf{x}$, $\mathbf{A}$, and $\mathbf{b}$ can be thought of as watermark in the inverse-feature domain, feature extraction matrix, and watermark in the feature domain, respectively. Since the constraints of (20) are all linear and the Hessian $\mathbf{H}$ is positive semidefinite, the objective function is a convex form and its solution is known to exist uniquely in the optimization theory. Thus, (20) can be solved through the simple convex quadratic programming [26]. This problem has a unique global minimum, and thus we can obtain the unique solution of this problem.

5. SIMULATION RESULTS

In order to evaluate the invisibility and robustness of the proposed algorithm, we take four H.263 videos: Foreman, Carphone, Mobile, and Paris, which are in the standard CIF format (352×288) with the frame-rate of 25 frame/s. We construct four scenes intentionally from the above video sequences in which the first 180 frames, 120 frames, 175 frames, and 125 frames from Foreman, Carphone, Mobile, and Paris are employed for tests, respectively. Watermark signals are embedded only in luminance for each scene. Also, we use MPEG-2 (704×480) sequences, Football (125 frames) and Flower Garden (85 frames), which have the frame-rate of 30 frame/s.

The robustness against incidental or intentional distortions can be measured by the correlation values. In the proposed scheme, two aspects should be considered; one is the positive detection ability in case that a watermark is present,

Table 1: Detection results for Foreman and Carphone sequences after H.263 compression.

QP	Foreman				Carphone			
	PSNR (dB)	Bit rate (kbps)	Compression ratio	Correlation	PSNR (dB)	Bit rate (kbps)	Compression ratio	Correlation
5	36.37	823.99	36.89 : 1	0.91	38.12	585.98	51.39 : 1	0.90
6	35.22	587.12	51.35 : 1	0.83	37.01	434.57	68.59 : 1	0.89
7	34.61	483.76	61.94 : 1	0.79	36.40	364.18	81.22 : 1	0.89
8	33.87	374.87	79.14 : 1	0.75	35.64	292.06	100.12 : 1	0.87
9	33.45	324.67	90.76 : 1	0.73	35.20	259.06	112.05 : 1	0.82
10	32.92	267.05	109.14 : 1	0.70	34.63	218.29	131.39 : 1	0.74
11	32.61	240.00	120.62 : 1	0.70	34.27	197.57	144.03 : 1	0.78
12	32.23	208.89	137.21 : 1	0.69	33.83	173.35	162.26 : 1	0.73
13	31.99	192.52	147.92 : 1	0.67	33.55	159.64	174.80 : 1	0.73
14	31.69	173.68	162.52 : 1	0.66	33.19	143.91	191.78 : 1	0.71

Table 2: Detection results for Mobile and Paris sequences after H.263 compression.

QP	Mobile				Paris			
	PSNR (dB)	Bit rate (kbps)	Compression ratio	Correlation	PSNR (dB)	Bit rate (kbps)	Compression ratio	Correlation
5	34.63	3707.53	8.19 : 1	0.87	35.94	897.60	33.15 : 1	0.91
6	32.78	2959.53	10.24 : 1	0.83	34.40	699.79	42.18 : 1	0.87
7	31.94	2517.63	12.01 : 1	0.72	33.53	592.19	49.51 : 1	0.86
8	30.70	2094.05	14.41 : 1	0.71	32.47	486.61	59.68 : 1	0.79
9	30.07	1838.34	16.39 : 1	0.65	31.90	424.33	67.91 : 1	0.80
10	29.15	1569.05	19.15 : 1	0.63	31.26	359.31	79.34 : 1	0.73
11	28.66	1405.17	21.34 : 1	0.61	30.95	321.43	87.97 : 1	0.70
12	27.95	1224.19	24.42 : 1	0.61	30.50	279.20	100.90 : 1	0.69
13	27.56	1109.03	26.89 : 1	0.59	30.24	254.43	108.90 : 1	0.69
14	26.96	980.76	30.31 : 1	0.57	29.86	224.45	121.88 : 1	0.67

in which the correlation values should be above a given threshold, and the other is the negative detection ability in case that a watermark is not present. In the computer simulation, various attacks, including video compression as well as intentional RST distortions, are applied to test the robustness. For these attacks, the overall performance may be evaluated by the relative difference between the correlation values when a watermark is present or not. As a result, the overall correlation value is compared with a threshold to determine whether the test video is watermarked. An experimental threshold is chosen to be 0.55, that is, a correlation value greater than or equal to 0.55 indicates the presence of the copyright information. A correlation value less than 0.55 indicates the absence of a watermark.

Due to the restricted transmission bandwidth or storage space, video data might suffer from a lossy compression. More specifically, video coding standards, such as MPEG-1/2/4 and H.26x, exploit the temporal and spatial correlations in the video sequence to achieve high compression ratio. We test the ability of the watermark to survive video coding for various compression rates. Each sequence is considered as a scene, where an identical watermark signal is embedded, and each watermarked scene is encoded with the H.263 or MPEG-2 coder. First, we employ the H.263 to encode CIF videos at the variable bit rate (VBR). That is, the H.263 coder with the fixed quantizers (QP = 5 ∼ 14) yields average bit rates from 823.99 to 173.68 kbps for Foreman, from 585.98 to 143.91 kbps for Carphone, from 3707.53 to 980.76 kbps for Mobile, and from 897.60 to 224.45 kbps for Paris, respectively, as shown in Tables 1 and 2. For the MPEG-2 sequences, the MPEG-2 coder encodes the two video scenes at the constant bit rate (CBR) from 8 Mbps to 2 Mbps, as shown in Table 3. The PSNR and bit rate results are varied according to the characteristics of each sequence. For example, a watermarked Foreman video frame encoded at 324.67 kbps is shown in Figure 5b, which has an objective quality of 33.45 dB on the average. However, Carphone sequence is encoded at 259.06 kbps with the same quantizer. Note that the Foreman sequence has a faster motion than the Carphone sequence, and as a result, it requires additional bit

Table 3: Detection results for Football and Flower Garden sequences after MPEG-2 compression.

Bit rate (Mbps)	Football			Flower Garden		
	PSNR (dB)	Compression ratio	Correlation	PSNR (dB)	Compression ratio	Correlation
8	34.63	15.19 : 1	0.87	31.29	15.19 : 1	0.88
6	32.96	20.24 : 1	0.82	29.38	20.24 : 1	0.75
4	30.61	30.31 : 1	0.70	26.86	30.31 : 1	0.69
2	26.43	61.05 : 1	0.65	23.34	60.97 : 1	0.59

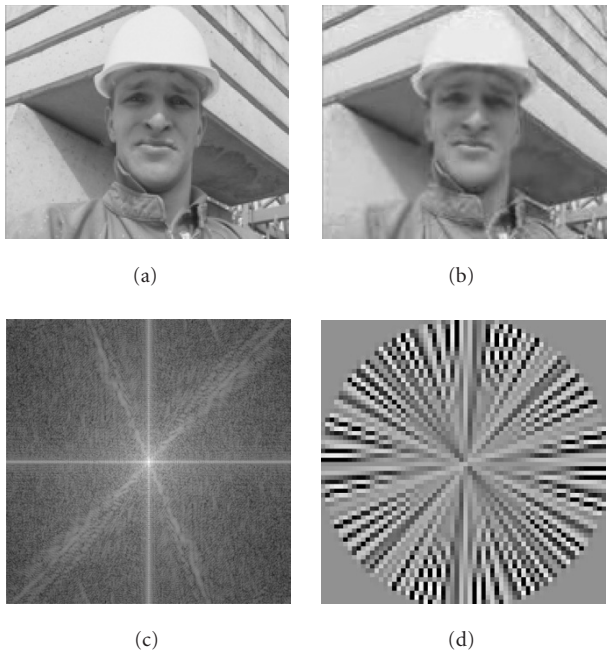


Figure 5: The 50th frame from Foreman sequence: (a) original, (b) watermarked and compressed, (c) 512 × 512 2D-DFT of (b), and (d) equalized watermark mask.

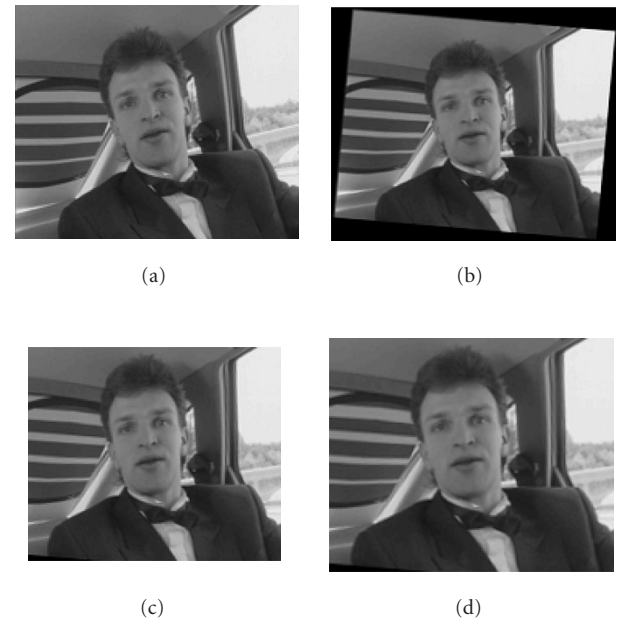


Figure 6: Examples of geometric attacks: (a) the original, (b) an image rotated by −5°, (c) a cropped image of (b), and (d) a resized image of (c) with the original image size.

rates to encode the video. Figure 5a is the original frame of Figure 5b. Figure 5c shows the 2D DFT magnitude of the watermarked frame in log-scale. The equalized watermark mask is shown in Figure 5d. As shown in Table 1, the watermarked Foreman sequence coded with compression ratio from 37:1 to 163:1 yields the detection results of correlation values from 0.91 to 0.66. Also, the results on the watermarked Carphone, Mobile, and Paris sequences are summarized in Tables 1 and 2 in which corresponding correlation values are from 0.90 to 0.71, from 0.87 to 0.57, and from 0.91 to 0.67, respectively. The detection results for the MPEG-2 video sequences are shown in Table 3. Each test is performed with 500 watermark keys. The detection results for the correct key are always above the given threshold 0.55, and the correlation values are under about 0.4 in case of no watermark.

Next, we illustrate the robustness of the proposed scheme against RST distortions. In most cases, RST distortions are accompanied by cropping. Figures 6a, 6b, 6c, and 6d show examples of rotation, rotation-cropping, and scaling for Carphone sequence, respectively. With the proposed algorithm, since the cropping does not lead to the loss of the synchronization, the disturbance from the cropping can be classified into the signal processing attacks. So, the distortion due to the cropping can be viewed as additive noise, which may degrade the detection value but not severely. In the simulation, each frame is modified with rotations of −5° and 5°, without or with cropping of maximum 16%, and scaling up to the original image size, as shown in Figure 6. Also, translation and scaling for each frame are performed.

The detection results after rotation without cropping for Foreman sequence are shown in Figure 7. Figure 7a shows the correlation values without rotation for 500 watermark keys, and Figures 7b and 7c show the correlation values after rotation by −5° and 5°, respectively. Figure 7d shows the detection results against rotation (−5° to 5°) without cropping, where the error bars indicate the maximum and minimum

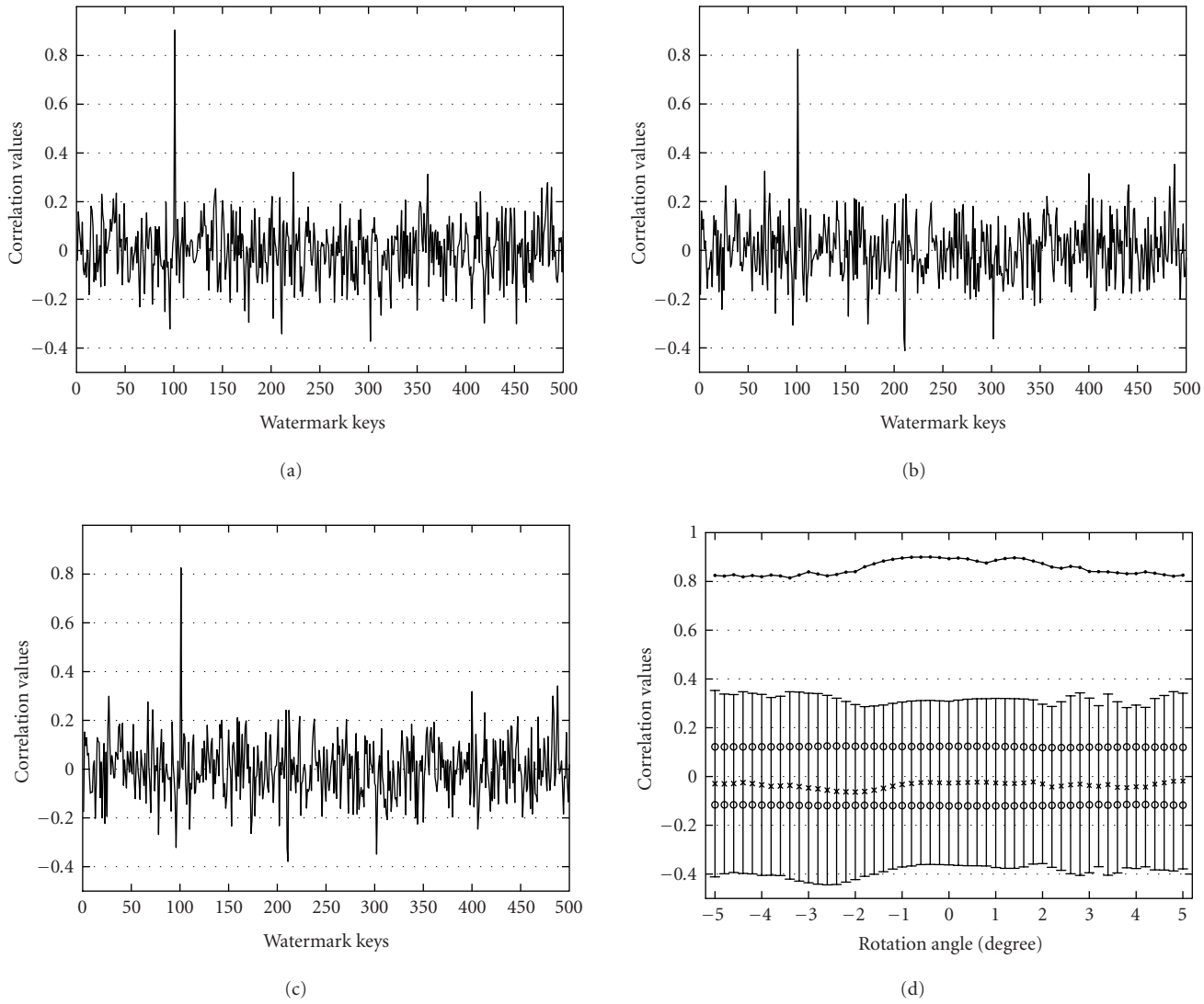


Figure 7: Correlation values after rotation without cropping for Foreman sequence: (a) detection without attacks, (b) detection after rotation by -5°, (c) detection after rotation by 5°, and (d) correlation values versus rotation angle without cropping.

correlation values over the 500 runs in case of no watermark. In Figures 8 and 9, the detection results after rotation without cropping for Carphone, Mobile, and Paris sequences are presented. The correlation values after rotation with cropping for various video sequences are shown in Figure 10. In all cases, the presence of a watermark is easily observed, and the maximum correlation values without watermark are under about 0.4. The DFT itself might be RST invariant, but it is often the case that the rotation with or without cropping yields noise-like distortions on the image. The simulation results show that these distortions affect correlation values only slightly in the proposed watermarking strategy.

The correlation detections on translation attacks are performed, and the plots are shown in Figure 11. In case of translation, we cropped the upper left part of each frame, and the reference position is translated, and the translation ratio in Figure 11 means noncropping ratio. Figure 12 shows the correlation values after scaling for various video sequences. Also, the presence of the embedded watermark is easily determined. Despite loss of 50 % or more by translation or scaling, the correlation results are maintained without much variance. In the proposed scheme, rotation and scaling in the frame domain yield a circular shift in the corresponding FPVs and decrease the power of them, respectively. They do not change the DFT magnitude of the PVTM, but the phase component only. As a result, in spite of noise-like distortions due to the RTS in the image domain, the WFV is almost invariant.

Some of the distortions of particular interest in video watermarking are those associated with temporal processing, for example, frame-rate change, temporal cropping, frame dropping, and frame interpolation. As usual, these uniform

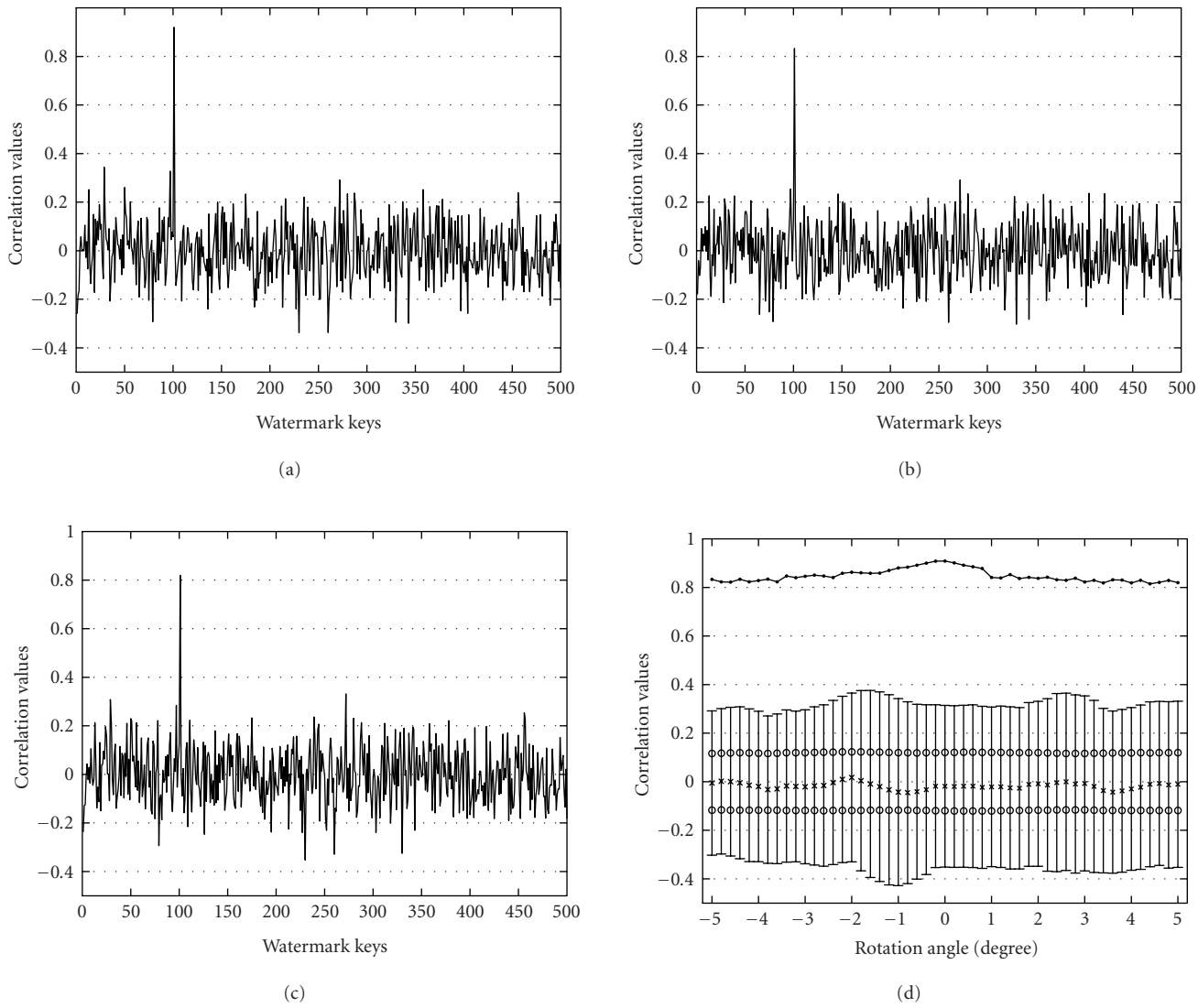


FIGURE 8: Correlation values after rotation without cropping for Carphone sequence: (a) detection without attacks, (b) detection after rotation by -5°, (c) detection after rotation by 5°, and (d) correlation values versus rotation angle without cropping.

temporal attacks may occur in common video processing, such as transcoding. To test frame dropping and interpolation, we dropped the odd index frames from the test sequences, which means that the frame-rate decreases to the half. For the case of frame averaging, the missing frames are replaced with the average of the two neighboring frames. In these cases, the proposed algorithm detects the watermark perfectly. In the proposed algorithm, the frame-rate change means scaling in the PVTM space. That is, most of the aforementioned temporal distortions are represented by those in the PVTM space. So, since these temporal attacks do not change the DFT magnitude of the PVTM, the WFV, extracted from the PVTM, is also invariant. The detection results after frame-rate changes, which can be achieved by dropping every nth frame, are shown in Figure 13 for various video sequences. Also, those after frame dropping and averaging are shown in Figure 14, in which every nth frame is interpolated by averaging its neighboring frames. As shown in Figures 13 and 14, it is shown that the detection value is reduced as the deformation increases, due to the frame averaging or frame dropping. This simulation shows, nevertheless, that the uniform frame-rate change cannot prevent the proposed algorithm from detecting the watermark signal even though half of the frames are lost. However, in case of random temporal attacks, that is, random frame dropping, it is expected that the proposed algorithm has a weakness not to find the synchronization and not to compensate the lost frames. Generally, the uniform deformation can be recovered by the DFT without loss of synchronization, which is shown by the proposed computer simulation. On

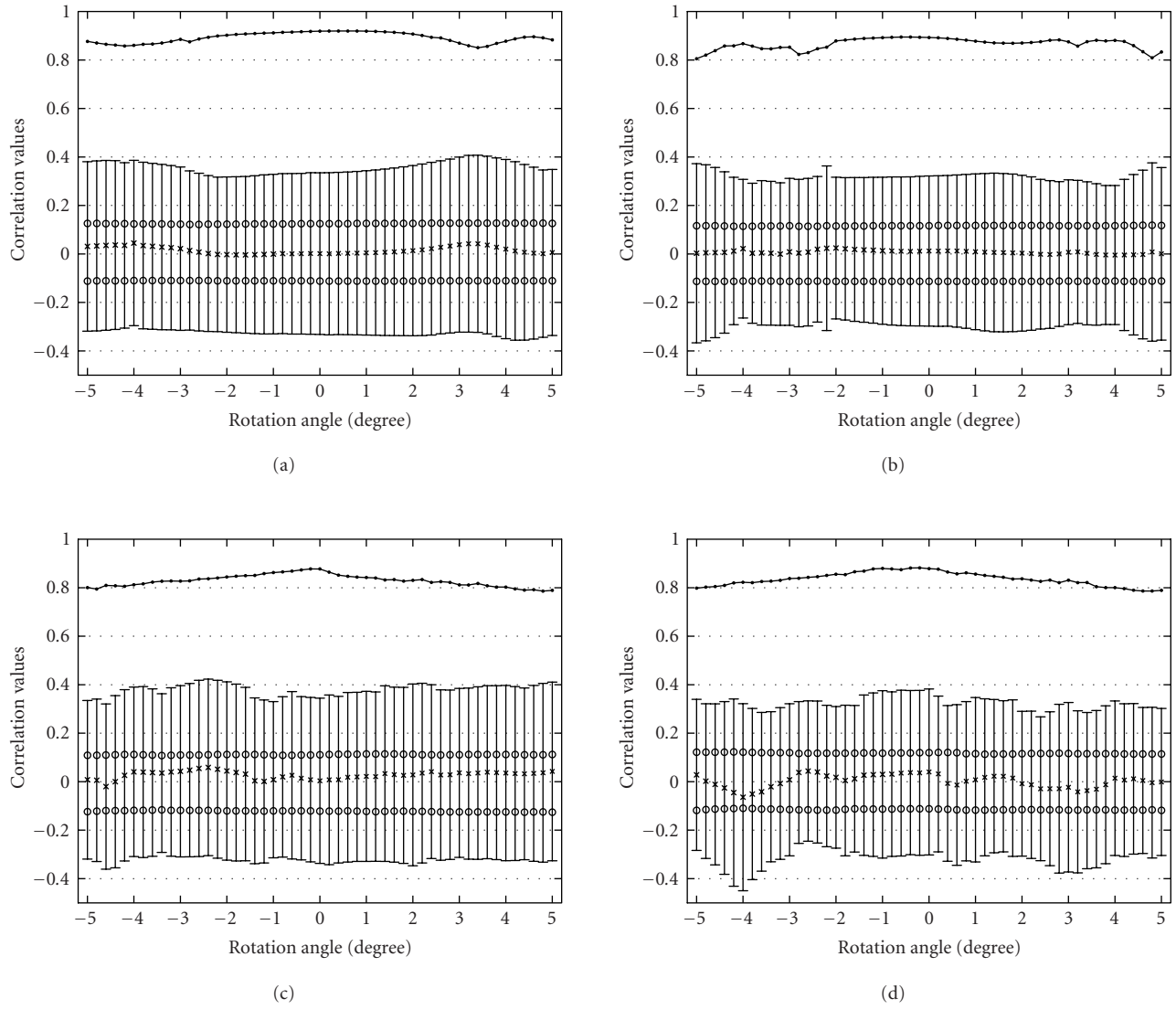


Figure 9: Correlation values after rotation without cropping for (a) Mobile, (b) Paris, (c) Football, and (d) Flower Garden sequences.

the other hand, random loss in the time or spatial domain must cause the DFT signals uncorrelated with the original. That is, the properties of the DFT cannot guarantee that the algorithm can recover the lost signals without exact information about their position. Thus, the proposed algorithm based on the DFT also cannot cope with the random frame dropping attacks and cover all general temporal attacks.

6. CONCLUSION

This paper presented a novel feature-based watermarking scheme for video sequences. In order to cope with video-oriented attacks, such as frame averaging, frame-rate changes, and interframe collusion, we employ a temporal watermarking algorithm, in which a watermark is embedded temporally to 1D projection vectors of the log-polar map, which is generated from the DFT of a 2D PVTM matrix. Each PVTM is segmented using well-known scene change detection algorithms. This strategy is very effective in order to cope with uniform temporal attacks. However, the proposed algorithm is not robust against random temporal attacks. In this paper, the feature extraction as well as its inverse processing were defined, and it was shown that the inverse problem yields a unique optimal solution subject to a few constraints. The computer simulation results demonstrated that the proposed scheme yields an acceptable performance for transparency and robustness against MC-DCT-based compression. Also, it was shown that the proposed scheme provides robustness to some video-oriented attacks, including frame-rate change, frame averaging, as well as interframe collusion.

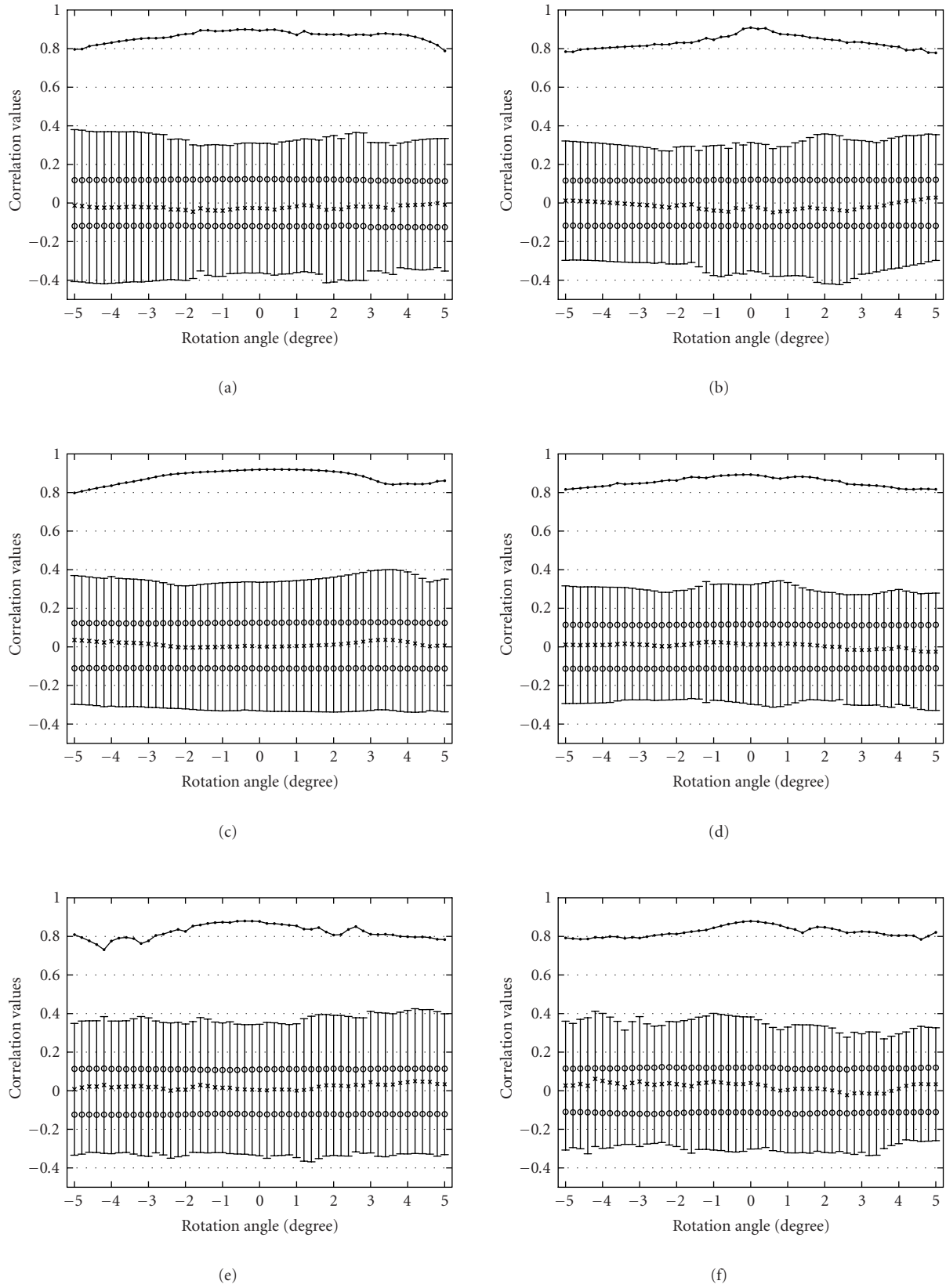


Figure 10: Correlation values after rotation with cropping for (a) Foreman, (b) Carphone, (c) Mobile, (d) Paris, (e) Football, and (f) Flower Garden sequences.

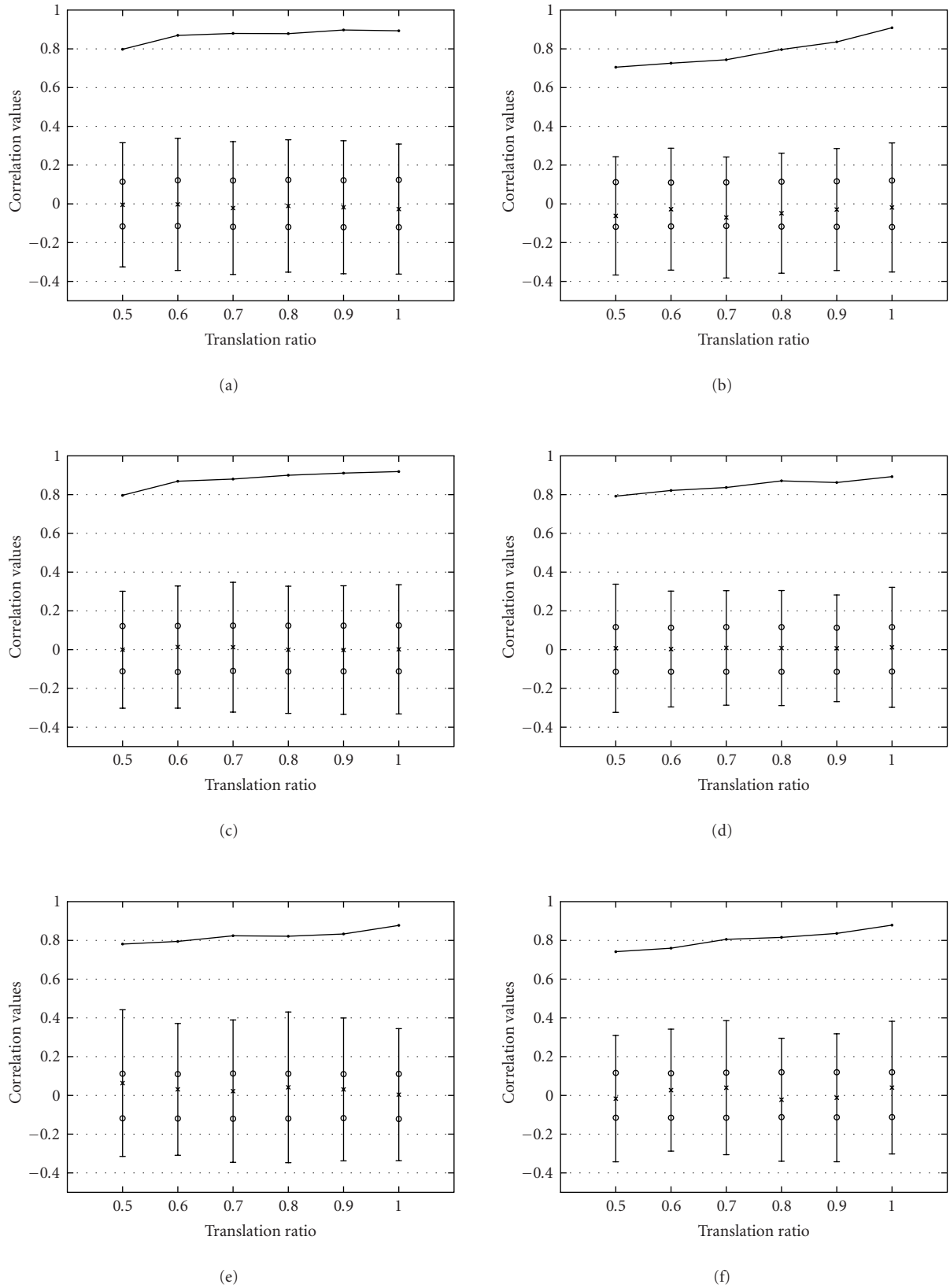


Figure 11: Correlation values after translation for (a) Foreman, (b) Carphone, (c) Mobile, (d) Paris, (e) Football, and (f) Flower Garden sequences.

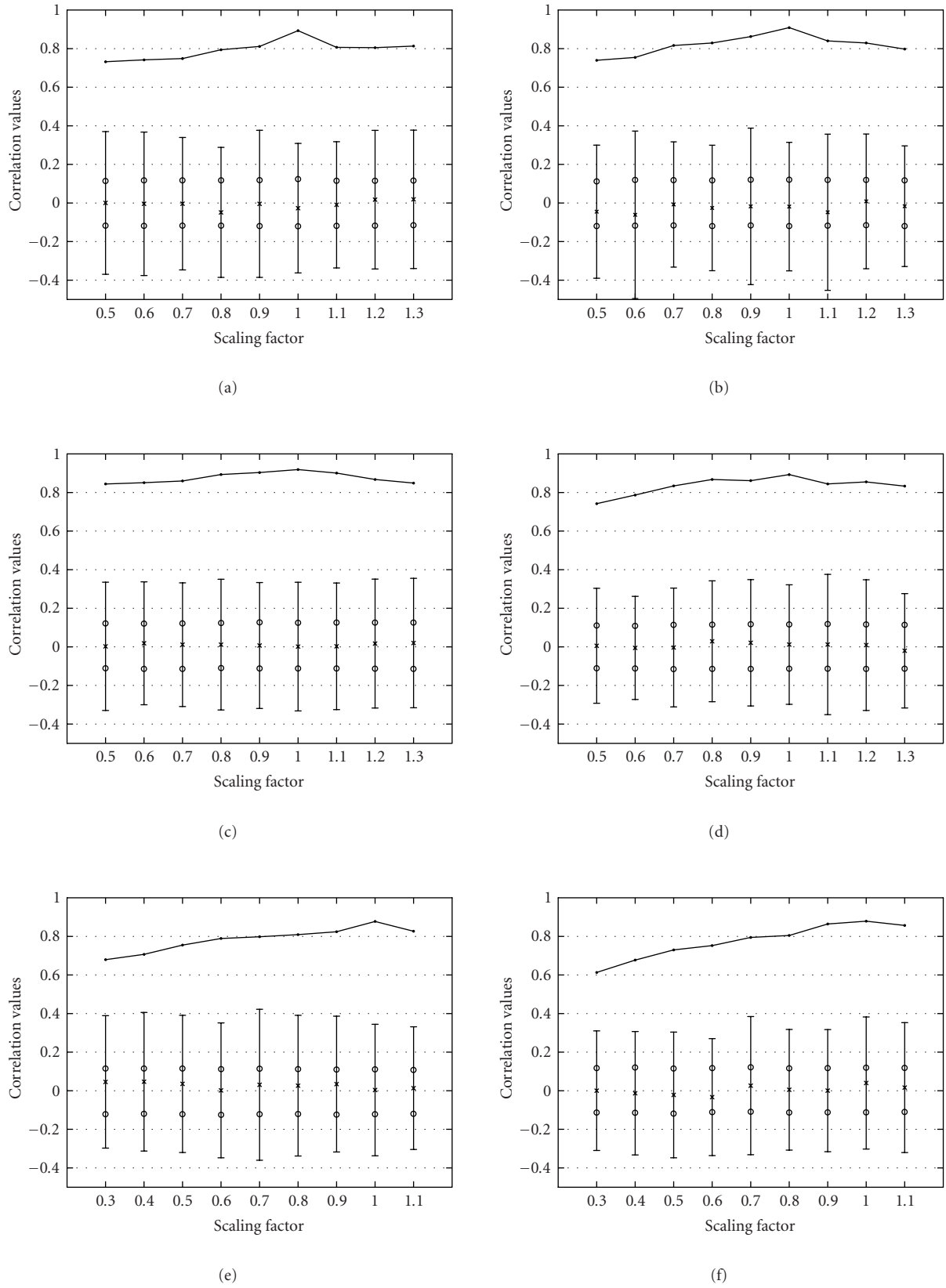


Figure 12: Correlation values versus scaling factor after scaling for (a) Foreman, (b) Carphone, (c) Mobile, (d) Paris, (e) Football, and (f) Flower Garden sequences.

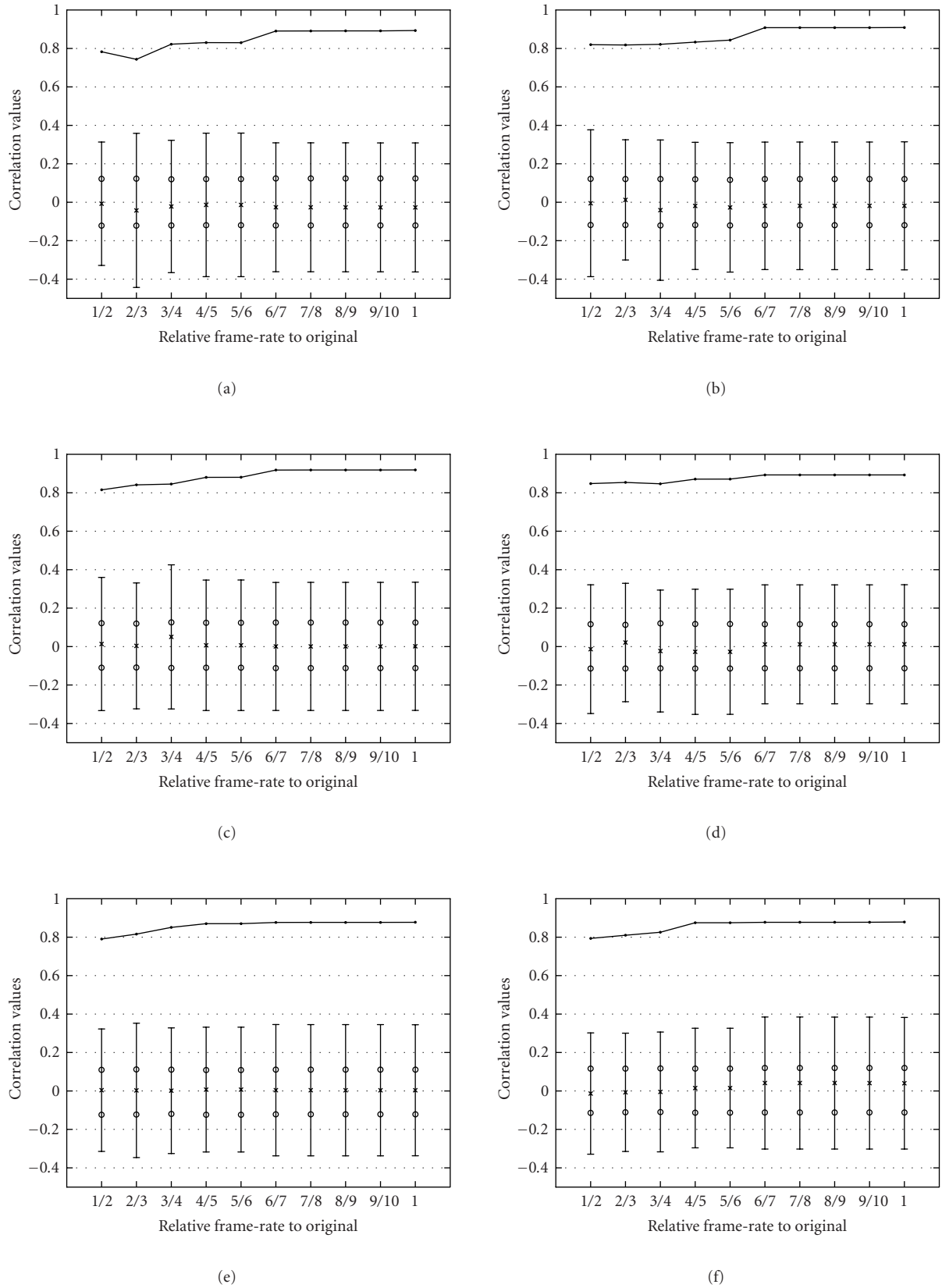


Figure 13: Correlation values after frame-rate changing by dropping for (a) Foreman, (b) Carphone, (c) Mobile, (d) Paris, (e) Football, and (f) Flower Garden sequences.

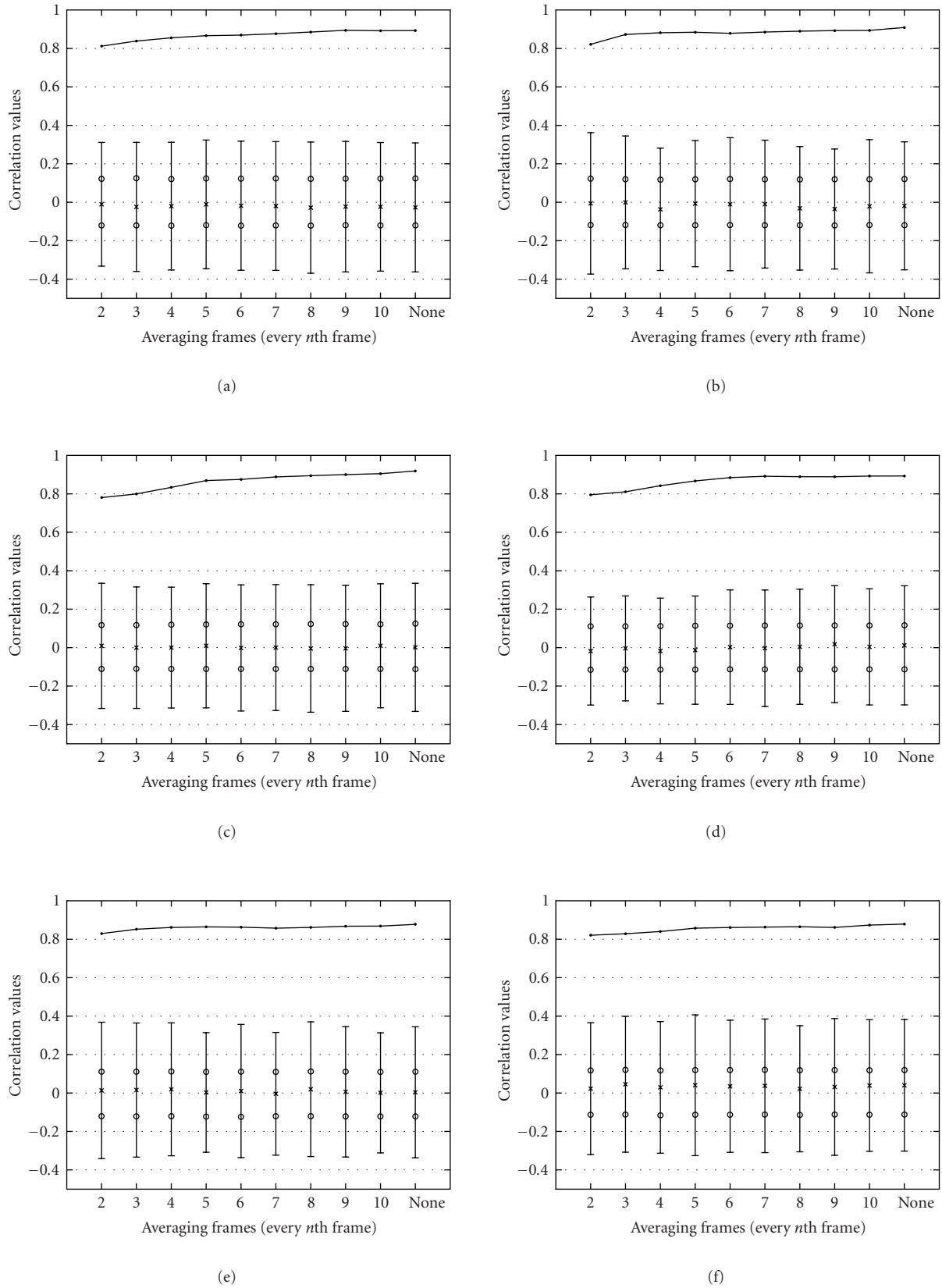


Figure 14: Correlation values after frame dropping and averaging for (a) Foreman, (b) Carphone, (c) Mobile, (d) Paris, (e) Football, and (f) Flower Garden sequences.

REFERENCES

[1] I. Pitas, "A method for signature casting on digital images," in *Proc. IEEE International Conference on Image Processing (ICIP '96)*, vol. 3, pp. 215–218, Lausanne, Switzerland, September 1996.

[2] R. B. Wolfgang and E. J. Delp, "A watermark for digital images," in *Proc. IEEE International Conference on Image Processing (ICIP '96)*, vol. 3, pp. 219–222, Lausanne, Switzerland, September 1996.

[3] G. C. Langelaar, J. C. A. van der Lubbe, and R. L. Lagendijk, "Robust labeling methods for copy protection of images," in *Storage and Retrieval for Image and Video Databases V*, vol. 3022 of *Proceedings of SPIE*, pp. 298–309, San Jose, Calif, USA, February 1997.

[4] I. J. Cox, M. L. Miller, and A. L. McKellips, "Watermarking as communications with side information," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1127–1141, 1999.

[5] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1673–1687, 1997.

[6] C. I. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 525–539, 1998.

[7] C.-Y. Lin, M. Wu, J. A. Bloom, I. J. Cox, M. L. Miller, and Y. M. Lui, "Rotation, scale, and translation resilient watermarking for images," *IEEE Trans. Image Processing*, vol. 10, no. 5, pp. 767–782, 2001.

[8] S. Pereira, J. J. K. O'Ruanaidh, F. Deguillaume, G. Csurka, and T. Pun, "Template based recovery of Fourier-based watermarks using log-polar and log-log maps," in *Proc. IEEE International Conference on Multimedia Computing and Systems (ICMCS '99)*, vol. 1, pp. 870–874, Florence, Italy, June 1999.

[9] J. J. K. O'Ruanaidh and T. Pun, "Rotation, scale and translation invariant spread spectrum digital image watermarking," *Signal Processing*, vol. 66, no. 3, pp. 303–317, 1998.

[10] J. Altmann and H. J. P. Reitbock, "A fast correlation method for scale- and translation-invariant pattern recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, no. 1, pp. 46–58, 1984.

[11] K. Su, D. Kundur, and D. Hatzinakos, "Novel approach to collusion-resistant video watermarking," in *Security and Watermarking of Multimedia Contents IV*, vol. 4675 of *Proceedings of SPIE*, pp. 491–502, San Jose, Calif, USA, January 2002.

[12] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Multiresolution scene-based video watermarking using perceptual models," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 540–550, 1998.

[13] F. Deguillaume, G. Csurka, J. J. K. O'Ruanaidh, and T. Pun, "Robust 3D DFT video watermarking," in *Security and Watermarking of Multimedia Contents*, vol. 3657 of *Proceedings of SPIE*, pp. 113–124, San Jose, Calif, USA, January 1999.

[14] W. Zhu, Z. Xiong, and Y.-Q. Zhang, "Multiresolution watermarking for images and video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 545–550, 1999.

[15] F. Hartung and B. Girod, "Digital watermarking of MPEG-2 coded video in the bitstream domain," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '97)*, vol. 4, pp. 2621–2624, Munich, Germany, April 1997.

[16] F. Hartung and B. Girod, "Watermarking of uncompressed and compressed video," *Signal Processing*, vol. 66, no. 3, pp. 283–301, 1998.

[17] G. C. Langelaar, R. L. Lagendijk, and J. Biemond, "Real-time labeling of MPEG-2 compressed video," *Journal of Visual Communication and Image Representation*, vol. 9, no. 4, pp. 256–270, 1998.

[18] G. C. Langelaar and R. L. Lagendijk, "Optimal differential energy watermarking of DCT encoded images and video," *IEEE Trans. Image Processing*, vol. 10, no. 1, pp. 148–158, 2001.

[19] M. M. Yeung and B. Liu, "Efficient matching and clustering of video shots," in *Proc. IEEE International Conference on Image Processing (ICIP '95)*, vol. 1, pp. 338–341, Washington, DC, USA, October 1995.

[20] H. S. Chang, S. Sull, and S. U. Lee, "Efficient video indexing scheme for content-based retrieval," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 8, pp. 1269–1279, 1999.

[21] G. Depovere, T. Kalker, and J.-P. Linnartz, "Improved watermark detection reliability using filtering before correlation," in *Proc. IEEE International Conference on Image Processing (ICIP '98)*, vol. 1, pp. 430–434, Chicago, Ill, USA, October 1998.

[22] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic, Boston, Mass, USA, 1992.

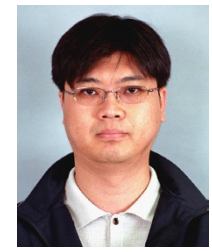
[23] S. Chatterjee, A. Hadi, and B. Price, *Regression Analysis by Example*, John Wiley & Sons, New York, NY, USA, 3rd edition, 2000.

[24] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1989.

[25] F. Goffin, J.-F. Delaigle, C. De Vleeschouwer, B. Macq, and J.-J. Quisquater, "Low-cost perceptive digital picture watermarking method," in *Storage and Retrieval for Image and Video Databases V*, vol. 3022 of *Proceedings of SPIE*, pp. 264–277, San Jose, Calif, USA, February 1997.

[26] R. J. Vanderbei, "LOQO: an interior point code for quadratic programming," *Optimization Methods and Software*, vol. 11, pp. 451–484, 1999.

[27] D. G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley, Reading, Mass, USA, 2nd edition, 1989.



Han-Seung Jung received the B.S., M.S., and Ph.D. degrees in electronic engineering from Seoul National University, Seoul, Korea, in 1996, 1998, and 2003, respectively. Since 2003, he has been with the Samsung Electronics Co., Suwon, Korea, working on the development of mobile application in the telecommunication network division. His current interests are in the areas of image and video signal processing, digital communication, and digital rights management system.

Young-Yoon Lee received the B.S. and M.S. degrees in electronic engineering from Seoul National University, Seoul, Korea, in 1999 and 2002, respectively. Currently, he is a Ph.D. candidate in electrical engineering at Seoul National University. His research interests include image processing, watermarking, and digital rights management.

Sang Uk Lee received the B.S. degree from Seoul National University, Seoul, Korea, in 1973, the M.S. degree from Iowa State University, Ames, in 1976, and the Ph.D. degree from the University of Southern California, Los Angeles, in 1980, all in electrical engineering. In 1980–1981, he was with the General Electric Company, Lynchburg, VA, working on the development of digital mobile radio. In 1981–1983, he was a member

of the technical staff, M/A-COM Research Center, Rockville, MD. In 1983, he joined the Department of Control and Instrumentation Engineering at Seoul National University as an Assistant Professor, where he is now a Professor in the School of Electrical Engineering. Currently, he is also affiliated with the Automation and System Research Institute and the Institute of New Media and Communications at Seoul National University. His current research interests are in the areas of image and video signal processing, digital communication, and computer vision. He served as an Editor-in-Chief for the *Transaction of the Korean Institute of Communication Science* from 1994 to 1996. Currently, he is a Member of the editorial board of the *Journal of Visual Communication and Image Representation* and an Associate Editor for IEEE Transactions on Circuits and Systems for Video Technology. He is a Member of Phi Kappa Phi.