Spread Spectrum Image Watermarking Based on Perceptual Quality Metric

Fan Zhang, *Member, IEEE*, Wenyu Liu, *Member, IEEE*, Weisi Lin, *Senior Member, IEEE*, King Ngi Ngan, *Fellow, IEEE*

Abstract— **Efficient image watermark calls for full exploitation of the perceptual distortion constraint. Second-order statistics of visual stimuli are regarded as critical features for perception. This paper proposes a second-order statistics (SOS) based image quality metric, which considers the texture masking effect and the contrast sensitivity in Karhunen-Loève transform domain. Compared with the state-of-the-art metrics, the quality prediction by SOS better correlates with several subjectively-rated image databases, in which the images are impaired by the typical coding and watermarking artifacts. With the explicit metric definition, spread spectrum watermarking is posed as an optimization problem: we search for a watermark to minimize the distortion of the watermarked image and to maximize the correlation between the watermark pattern and the spread spectrum carrier. The simple metric guarantee the optimal watermark a closed form solution and a fast implementation. The experiments show that the proposed watermarking scheme can take full advantage of the distortion constraint and improve the robustness in return.**

***Index Terms*— image quality metric, perceptual watermark, subjectively-rated database**

I. INTRODUCTION

Image watermarking is applied in copyright management (copyright authentication, traitor tracing, and access control) as well as channel error resilience. Perceptual watermarking (except fragile watermarking) exploits aspects of the human visual system (HVS) and seeks an invisible and robust watermark. This process can be generally formulated as:

$$\min_{Y}\{D(X,Y)-\gamma R(X,Y;m)\} \quad (1),$$

where $D(X, Y)$ is the perceptual distortion of the watermarked image Y with respect to the cover image X; R evaluates the robustness of the watermark. For instance, conventionally in spread spectrum watermarking scheme, the secret message m may be compressed to reduce the data load, then utilizing a key, a pseudo-random sequence is generated to modulate the secret message and added to the cover image. To take into account the perceptual impact, γ is a positive multiplier to balance the invisibility and the robustness, since robust watermark may involve with a high watermark strength and thus causes large distortion.

Some work of perceptual watermarking does not rely on an explicit image quality metric. For example, in [1-3] the upper bounds of watermark strength is set by the Just Noticeable Difference (JND) [4], which indicates the extent of a signal difference pattern being just perceived. These methods were usually testified by the subjective test on whether the watermark was visible or not with a given watermark strength. But the critical question is whether the watermark takes full advantage of the perceptual distortion constraint and achieves the robustness as high as possible. However, because there is no explicit metric, it is difficult to justify: (1) whether the watermark invisibility is accurately evaluated; (2) whether the watermark takes full account of the relevant HVS characteristics; (3) whether such heuristic embedding strength is the optimal for the tradeoff between the invisibility and the robustness of watermark. This study is devoted to perceptual watermarking with an explicit distortion metric whose accuracy is to be quantitatively evaluated by the subjectively-rated databases (benchmarking for image quality assessment), and meanwhile, watermarking can be posed as an optimization problem with an analytic solution.

To make Problem (1) easy to optimize, we need to keep the metric D simple. Note that for spread spectrum watermark, R is a linear function with respect to Y as to be explained in Sec V, a good metric is able to ensure Problem (1) with an analytic solution. For instance, if D is differentiable with respect to Y, Problem (1) can be approached by any gradient-based technique; if D is absolute convex with Y, Problem (1) may have a closed-form solution and the watermarked image can be computed without iterations. Fortunately, many image quality metrics can be generalized to a form below, which makes Problem (1) easy to be solved:

$$D(X,Y)=\left\|\mathbf{MH}(\vec{X}-\vec{Y})\right\|_{\ell} \quad (2),$$

where the cover image and the watermarked image are respectively vectorized in the raster scanning order to column vector $\vec{X}$, $\vec{Y}\in\Re^{q}$, and q is the number of pixels. According to the classification in [5], many metrics including [6-8] conform to Definition (2) and belong to the error sensitivity based metrics. In those metrics, image difference is calculated at first, and then is decomposed into multiple channels, which are related to the neural responses in the HVS. This channel

Manuscript received October 14, 2010; revised January 30, 2011 and April 18, 2011. This work is partially supported by NSFC, China (Grant No.: 60903172, 60873127) and MoE AcRF Tire 2 Grant, Singapore (Grant No.: T208B1218). This paper was recommended by Associated Editor J. Dittmann.

Fan Zhang and Wenyu Liu are with Huazhong University of Science and Technology, China (e-mail: fan.zhang@technicolor. com, liuwy@hust.edu.cn); Weisi Lin is with Nanyang Technological University, Singapore (e-mail: wslin@ntu.edu.sg); King Ngi Ngan is with the Chinese University of Hong Kong (e-mail: knngan@ee.cuhk.edu.hk).

decomposition is performed by $\mathbf{H} \in \Re^{p \times q}$ that may be selective for spatial location, frequency band and orientation. For instance, **H** can be a Toeplitz matrix if a 2D convolution operator (e.g. a Gabor filter) is used to decompose the HVS channel [9], and alternatively, **H** can also represent any orthogonal transform based on non-overlapping blocks. In these two cases, p is equal to q. If **H** is mapped from an over-completed transform, p is bigger than q. Finally, the image difference in each channel is weighted by the diagonal matrix $\mathbf{M} \in \Re^{p \times p}$. Several HVS related features may be encapsulated into the weighting matrix **M**:

- *Contrast Sensitivity Function (CSF):* it models the sensitivity of the HVS as a function of the spatial frequency in visual stimuli. It usually peaks at the middle frequency and drops with both increasing and decreasing frequencies. The CSF can be reflected in **H** with Fourier transform [6], DCT [7], and wavelet transform [8].
- *Texture masking effect* also termed as noise visibility function. It refers to the reduction of visibility of noise-like watermark due to the presence of the cover image which shows similar frequency and orientation content. Generally, the watermark is more difficult to be noticed in the texture region.
- *Luminance masking effect* also termed as light adaption. It indicates the watermark is less visible in the very dark or bright region of the cover image.

Finally, the differences from all the channels are pooled into a scalar score by the ℓ-order Minkowski norm (denoted by $\|.\|_l$). MSE (Mean Squared Error) is the simplest case of Definition (2), where **H** and **M** are both the identity matrices $\mathbf{I}^{q \times q}$. SNR (Signal to Noise Ratio) and PSNR (Peask Signal to Noise Ratio) are monotonic regression from the MSE and thus are equivalent to MSE. Definition (2) is differentiable with Y, and more specifically, when 2nd-order Minkowski norm is used, Definition (2) makes Problem (1) have an analytic solution as explained in Sec V.

Definition (2) keeps concise due to ignoring some nonlinear characteristics, for example suprathreshold [10] and threshold effect [4]. Suparathreshold effect refers to the fact that the same amount of distortion becomes perceptually less significant when the overall distortion level increases, which seldom happens in watermarking because watermarks often avoid causing heavy distortion. Threshold effect refers to the invisibility of distortions below the JND, and need to be captured by a piece-wise function, which is indifferentiable. Definition (2) does not cover a few metrics, like Visual Information Fidelity (VIF) [11] and Structural Similarity Index (SSIM) [5]. Perceptual image processing by optimizing SSIM is nontrivial work yet [12]. No literature has reported how to guide image processing by optimizing VIF. To conclude, Definition (2) provides a metric framework, which embodies the typical HVS characteristics and makes watermarking problem easy to optimize.

In this paper, our metric design conforms to Definition (2). That is, we focus on the selection of **H**, **M** and l. Traditionally, the CSF associated with the image independent transform (e.g. DCT or wavelet) is employed, where the frequency subbands correspond to fixed visual patterns and the CSF parameters have to be estimated by visualpsychological experiments. Such experiments are sensitive to the viewing conditions. This paper uses the Karhunen-Loève transform to decompose the frequency subbands and the corresponding eigenvalue to calculate the CSF parameters. Without tuning the parameters, the proposed metric achieves a better overall performance than the state-of-the-art metrics on subjectively-rated databases, and meanwhile the optimal spread spectrum watermark has a closed form solution and can be computed without iterations.

The novelty of this paper includes the following aspects. (1) The spread spectrum watermarking is posed as an optimization problem with simplified formulation of the invisibility and the robustness. (2) The proposed quality metric is simple but outperforms the state-of-the-art metrics in overall, and ensures watermarking problem an analytic solution. (3) Under the guidance of the proposed metric, the watermark can be efficiently detected.

It is important to clarify the watermark invisibility. Invisibility generally means that the watermark cannot be perceived. However, the "visible" watermark sometimes is also acceptable when it is not annoying (i.e., without degrading the image quality too much) or is hardly distinguished from the passive channel noises. Some applications pay more attention to the watermark payload and robustness rather than the distortion. In a word, watermarking is a tradeoff. Keeping this in mind, the quality of watermarked image with respect to the cover image is linked with the *loosened invisibility* (i.e. relaxed distortion constraint) of watermark.

The rest of this paper is organized as follows. After a brief description of the related metric for watermarking in Sec. II, a new metric is proposed in Sec. III. The metric performance is compared in Sec. IV with seven relevant existing metrics on six publicly accessible databases. The corresponding watermarking scheme is proposed in Sec. V. Sec. VI is devoted to comparison of the watermark performance. A smart detector associated with the proposed embedding method is presented and evaluated in Sec. VII, followed by the conclusion in Sec. VIII.

II. Related works

Image quality assessment is a common problem for many applications including image restoration, coding as well as watermarking. Image quality metric can be classified into full reference (FR), reduced reference and no reference, according to the availability of the distortion free image (i.e. the cover image in watermarking), which may be used as the reference to evaluate the distorted counterpart. The FR metric can provide the more reliable assessment concerning the annoyance introduced by watermark. PSNR and WPSNR (Weighted PSNR) [13] are currently the most commonly used for watermarking, due to their simplicity. It is well acknowledged that PSNR does not predict the image quality well. WPSNR is more accurate due to consideration of the texture masking effect; yet it does not consider the CSF and thus not helpful for watermarking in the frequency domain, since the watermark is usually perceptually shaped in the frequency domain according to the CSF. Many other FR metrics have high correlation with subjectivities, e.g. SSIM, VIF. However, they are not specially designed for watermarking and make Problem (1) difficult to optimize. Most

of them can only post-evaluate the watermark output, but cannot constrain the watermark in loop. The accuracy of the typical FR metrics from the viewpoint of watermarking is compared in [14], where a new metric CPA is also proposed, yet how to apply the CPA metric to watermarking is still a problem.

In this paper, we pay special attention to two typical metrics: the metric implied in [15] (we name it as LSE metric for it is based on the Least-Squares prediction Error within the cover image) and DCTune [7]. Both metrics have been incorporated in watermarking. LSE is a recent method with only consideration of the texture masking effect, similar to the WPSNR and several proposals in [16]. DCTune is a typical JND based metric. It was originally developed for perceptual coding, and widely used in watermarking schemes [1, 17] and adopted to evaluate the quality of watermarked images in the Checkmark Benchmarking Tool [18]. Many perceptual metrics in various spatial frequency domains can be thought as the variations of the DCTune. We briefly introduce these two metrics and their conformity with Definition (2).

- *LSE metric*

The LSE metric estimates the texture masking effect from the least-squares (LS) prediction error within the cover image. The absolute error of filtering prediction is defined as

$$M = \left| X - g * X \right| \tag{3}.$$

In Eq. (3), $g * X$ represents convoluting (noncausally filtering) the image $X \in \Re^{I \times J}$ by an LS filter $g \in \Re^{5\times 5}$, where image X is supposed to have I rows and J columns, i.e., $q = I \cdot J$; M is of the same size with X and has nonnegative elements. Filter g is learned by LS prediction: $g = \arg\min \| X - g*X \|^2$, and is the filter to minimize the squared prediction error. In [15], the LSE watermark strength is proportional to M. This heuristic scheme is actually the optimal solution to the problem (1) when the robustness is commonly defined (like (14) in Sec. II) and the distortion is defined by

$$D(X,Y) = \frac{1}{q}\sum_{i,j}\frac{(x_{i,j} - y_{i,j})^2}{M_{i,j} + b} \tag{4},$$

where $x_{i,j}$ and $y_{i,j}$ are the pixel intensity of image $X \in \Re^{I\times J}$ and $Y \in \Re^{I\times J}$ at the location (i, j) respectively ($1 \le i \le I$, $1 \le j \le J$); since $M_{i,j}$ is nonnegative, the LSE watermark scheme [15] embeds strong watermarks in the rough region with large $M_{i,j}$, slight watermarks in the smooth region with small $M_{i,j}$, and specifically no watermark in the region with $M_{i,j} = 0$. To ensure that the LSE metric, i.e. Definition (4), returns a limited value, we need set b as a positive constant to avoid dividing by a zero.

The LSE metric is a special case of Definition (2), considering that $\mathbf{H}$ is replaced by the identity matrix $\mathbf{I}^{q\times q}$, the diagonal elements of $\mathbf{M} \in \Re^{q\times q}$ consist of the reciprocal square root of $M_{i,j} + b$, and the 2nd order norm is adopted.

- *DCTune*

DCTune is defined as

$$D(X,Y) = \left\{ \frac{1}{q}\sum_{i,j,k}\frac{1}{M_{i,j,k}}(u_{i,j,k} - v_{i,j,k})^4 \right\}^{1/4} \tag{5},$$

$$M_{i,j,k} = \max\left\{ t_{i,j}\left(\frac{u_{0,0,k}}{\bar{u}_{0,0}}\right)^{0.694},\ \left|u_{i,j,k}\right|^{0.7}\left(t_{i,j}\left(\frac{u_{0,0,k}}{\bar{u}_{0,0}}\right)^{0.694}\right)^{0.3} \right\}$$

where i and j index the frequency band of 8×8 DCT transform ($0 \le i \le 7$, $0 \le j \le 7$), and k indexes the block (suppose a total of K blocks, i.e. $k = 1, 2, \ldots, K$ and $q = 8\times 8\times K$); u and v are the DCT coefficients of image X and Y respectively; $u_{0,0,k}$ is the DC coefficient of k-th block and $\bar{u}_{0,0}$ is the mean of all $u_{0,0,k}$. DCTune is a special case of Definition (2), where the 4th order norm is used, $\mathbf{H}$ is mapped from DCT and the diagonal element of $\mathbf{M}$ consist of the reciprocal square root of $M_{i,j,k}$. In the definition of M, constant $t_{i,j}$ is the CSF, $u_{i,j,k}$ plays the role of the texture masking effect, $u_{0,0,k}$ functions as the luminance masking effect, and the parameters are determined from psychophysical experiments.

A recent work related to ours was in [19], where the distortion constraint on the watermark consists of two separated ones: the CSF based one and the texture-masking based one. Such separation has three drawbacks. First, it burdens the distortion controlling with an additional balance among the separated constraints. Secondly, it is difficult to evaluate the accuracy of the separated distortion definition. Finally, it has to have multisteps to compute the optimal watermark. In this paper, both the CSF and the texture masking effect are encapsulated into a single explicit metric, so its performance can be tested and a good watermark shaping can be easily computed.

Our approach is also different from CPA [14] and the distortion definition in [19] in that we use an efficient CSF which comes from the statistics of the cover image but do not rely on actual HVS model parameters. Our metric is therefore not sensitive to the displaying characteristics or the viewing configurations.

III. Proposed Metric

Second-order statistics are regarded as critical features for visual pattern discrimination [20]. Karhunen-Lo ève transform is a projection whose directions are associated with the second-order statistics of the data samples. Along the principal directions, samples present a large variance (the second-order moment). Such directions capture the typical appearance of the samples and are often used for matching. The local variance (i.e. the luminance variance of the local image block) is another second-order statistics, which has been exhaustively studied and usually employed to capture the texture masking effect in images. The proposed metric is based upon the Second-Order Statistics based metric (SOS).

If the image is divided into $a \times a$ nonoverlapping blocks and the k-th block of the cover image and the watermarked image are vectorized and denoted by $\vec{x}_k$ and $\vec{y}_k$ respectively, the SOS metric is defined as

$$D(X,Y) = \left(\sum_k \frac{(\vec{x}_k - \vec{y}_k)^{\mathrm{T}}\mathbf{S}(\vec{x}_k - \vec{y}_k)}{(\sigma_k + h)\cdot T} \right)^{1/2} \tag{6},$$

where σ_k is the standard deviation of image block x_k, h is a constant for avoiding dividing by zero, $\mathbf{S}$ is the covariance

matrix about all the vectors $\vec{x}$ within the cover image, and T is the trace of **S**, i.e., the sum of all the eigenvalues of **S**.

By the eigendecomposition of the covariance matrix, we have $\mathbf{S} = \mathbf{U}\ \mathbf{\Lambda}\ \mathbf{U}^{\mathrm{T}}$. Generally, the covariance matrix **S** derived from natural images is nonsingular, so $\mathbf{U}^{\mathrm{T}}$ represents the complete KLT, and $\mathbf{\Lambda}$ has a^2 eigenvalues, $\lambda_1\ \lambda_2\ \ldots\ \lambda_{a^2}$, as its diagonal elements. Definition (6) can be rewritten in the KLT domain, after absorbing T ($T = \lambda_1 + \lambda_2 + \ldots + \lambda_{a^2}$) into $\mathbf{\Lambda}$:

$$D(X,Y) = \frac{1}{q}\sum_{k=1}^{K}\left(\frac{1}{(\sigma_k + h)}\sum_{i=1}^{a^2}\hat{\lambda}_i (u_{i,k} - v_{i,k})^2\right) \qquad (7),$$

where $\hat{\lambda}_i$ represents the eigenvalue normalized by T, $u_{i,k}$ and $v_{i,k}$ are the i-th KLT coefficients of block x_k and y_k respectively, and $q = k \times a^2$. Definition (7) is a nested summation; the inner summation pools the errors in the KLT subbands while the outer summation pools errors spatially. $\hat{\lambda}_i$ are the weights of the inner summation and plays a similar role as the CSF. Here, the CSF is considered in the KLT domain. As well known, KLT is more likely a kind of spatial-frequency channel decomposition. In [21], it is explained why the principal components of natural image patches are always two-dimensional sinusoids when recovered to the patch case, that is, why the KLT subbands are closely related to the oscillation frequency. Natural images often have the energy concentrated in their low frequency sub-bands, and the larger eigenvalues just correspond to the image energy in the lower and middle frequency sub-bands. With the CSF of $\hat{\lambda}_i$, the metric emphasizes the distortion in the low frequency bands while tolerates more distortion in the high frequency bands. On the other hand, the reciprocal of $(\sigma_k + h)$ is the weight of the outer summation and represents the texture masking. The metric thus emphasizes the distortion in the smooth region while tolerating the distortion in textures and edges. The SOS metric follows Definition (2), considering that **H** is mapped from KLT and the diagonal elements of **M** consist of $\sqrt{\hat{\lambda}_i / (\sigma_k + h)}$.

It is interesting to compare the proposed metric with Mahalanobis distance [22] as defined in (8), which is a dissimilarity metric between two random vectors $\vec{x}$ and $\vec{y}$ of the same distribution with the covariance matrix **S**. A notable difference between Mahalanobis distance and the proposed metric is whether $\hat{\lambda}_i$ or the reciprocal $\hat{\lambda}_i$ is used to weigh the distance along the eigenvector. In other words, should the distortion along the principal components be emphasized or be tolerated? Mahalanobis distance has an intuitive explanation: when a test point seems to belong to a mass of sample points, it must be close to the mass center along the non-principal axis where the sample points spread out over a small range, but could be a little far along the principal axis where the sample points spread out over a large range. Obviously, Mahalanobis distance tolerates the distortions along the principal components while the proposed metric emphasizes them. Although Mahalanobis distance is often used to detect outliers in statistical testing, we find that it is not suitable for measuring the quality of natural images. The reason is probably that the HVS has been evolved to pay more attention to the principal components of images, where images show more diversity and uncertainty. In this way, humans can efficiently learn from natural images. Therefore, the distortion along the principal components should be emphasized rather than be tolerated. In other words, the CSF of $\hat{\lambda}_i$ can be explained as the HVS's adaption to the prior of stimuli. Note that the proposed CSF does not rely on actual HVS model parameters and just is the statistics of the cover image.

$$\begin{aligned} D(x_k, y_k) &= \left((\vec{x}_k - \vec{y}_k)^{\mathrm{T}}\mathbf{S}^{-1}(\vec{x}_k - \vec{y}_k)\right)^{1/2} \\ &= \left(\mathbf{U}^{\mathrm{T}}(\vec{x}_k - \vec{y}_k)^{\mathrm{T}}\mathbf{\Lambda}^{-1}\mathbf{U}(\vec{x}_k - \vec{y}_k)\right)^{1/2} \\ &= \left(\sum_{i=1}^{a^2}\frac{(u_{i,k} - v_{i,k})^2}{\hat{\lambda}_i}\right)^{1/2} \end{aligned} \qquad (8).$$

IV. Metric Performance

It is important to make sure that the proposed metric works well before applying it to watermarking, so this section focuses on comparison of the metric performance from the viewpoint of watermarking. The existing benchmark tools for watermarking often use PSNR or WPSNR to evaluate the quality of watermarked images but pay little attention to justify whether the metrics correlates well with the subjectivities in quality evaluation. [23]. On the contrary, several subjectively-rated databases are built and can be taken as the "ground truth" for evaluating the metrics. The databases contain the reference (or cover) images, the impaired (or watermarked) images and the corresponding subjective quality scores rated by human observers. The subjective assessment tests generally follow the procedures under normalized viewing conditions as those defined by the ITU (International Telecommunications Union) recommendations, e.g. [24]. The accuracy of a metric is evaluated by the correlation between its objective scores and the subjective scores of the database assigned to each impaired images. Specifically, the LCC (Linear Correlation Coefficient or Pearson correlation) and SROCC (Spearman Rank Order Correlation Coefficient) are used to assess how well the relationship between the subjective and the objective scores can be described using a monotonic function. Calculating LCC requires that the objective scores are monotonic regressed by the logistic function proposed in [25]; while SROCC does not depends on the monotonic regression. The larger LCC and SROCC values mean the higher correlation. The best LCC and SROCC of 1 will occur when the objective scores is the perfect monotonic function of the subjective ones. The reader may find that Definitions (4) and (7) are not the strict 2nd-order norm due to lack of square root. However, the omission does not affect the LCC and SROCC performance of the metric, since the square root is also a monotonic function and will be compensated by the regression.

The early subjectively-rated databases were designed for coding applications [26-28], while the databases for watermarking were also developed recently [29-31]. A main difference between them is the range of distortion levels. Severe distortions are often covered by the databases for coding but not contained those for watermarking due to the distortion constraint on watermark. Another difference is the types of image distortions. The databases for coding frequently take into account the compression, additive noise, blur, network

transmission error, etc., while the databases for watermarking collect the typical watermarking algorithms to generate distortions. In practice, the watermarks may consist of the distortions in coding applications but are more likely to have more diverse frequency spectrum. It is worth testing the metrics for watermarking by the databases for coding, since there are no fixed formularies for watermarks and any perceptually shaped noises are possible choices for spread spectrum watermark. We highlight the importance of using sufficient databases to evaluate metrics. Due to the analysis above, the comparative experiment is configured below.

TABLE I
METRIC PERFORMANCE ON SUBJECTIVELY-RATED DATABASES

		FourierSB	BA	Meerwald	LIVE	TID	CSIQ	LIVE half set	TID half set	CSIQ half set
Number of images		210	120	120	779	1400	750	389	700	375
VIF	LCC	0.887	0.945	0.895	**0.960**	**0.887**	0.931	**0.851**	0.642	0.770
	SROCC	0.862	0.934	0.890	**0.964**	0.864	0.928	**0.850**	0.632	0.778
MSSIM	LCC	0.907	0.900	0.929	0.942	0.856	**0.948**	0.820	0.532	**0.839**
	SROCC	0.865	0.896	0.919	0.945	**0.866**	**0.950**	0.819	0.553	0.830
PSNR	LCC	0.774	0.957	0.918	0.872	0.734	0.910	0.743	0.495	0.784
	SROCC	0.696	0.934	0.891	0.856	0.740	0.906	0.774	0.490	0.781
WPSNR	LCC	0.793	0.954	0.938	0.903	0.746	0.907	0.469	0.630	0.797
	SROCC	0.707	0.949	0.928	0.909	0.782	0.915	0.481	0.628	0.794
CPA	LCC	**0.914**	**0.968**	**0.954**	0.864	0.683	0.945	0.735	0.573	0.824
	SROCC	**0.919**	**0.962**	**0.949**	0.868	0.682	0.946	0.702	0.575	0.822
LSE	LCC	0.784	0.956	0.930	0.887	0.758	0.915	0.806	0.577	0.808
	SROCC	0.708	0.947	0.925	0.891	0.779	0.918	0.802	0.562	0.800
DCTune	LCC	0.726	0.905	0.871	0.820	0.711	0.750	0.772	0.638	0.671
	SROCC	0.659	0.882	0.857	0.823	0.746	0.766	0.768	0.628	0.682
SOS	LCC	0.895	0.951	0.931	0.940	0.851	0.945	0.838	**0.647**	0.830
	SROCC	0.883	0.946	0.928	0.944	0.865	0.944	0.834	**0.642**	**0.831**

- *Databases*

Six databases, including three databases for watermarking and three databases for coding are selected. Databases of FourierSB [29], BA [30] and Meerwald [31] have the different watermarks as distortions. A short description about them is contained in [14]. Databases of LIVE [26], TID [32] and CSIQ [28] have diversified distortion types for coding applications and sufficient distorted images. We exclude the distortions of "local block-wise distortions of different intensity" and "mean shift" in TID, as well as "contrast change" in TID and CSIQ. These distortions are difficult to handle by most metrics and seldom happen during embedding spread spectrum watermarks. From each one of LIVE, TID, and CSIQ, we also choose the image subset with the highest half of quality as a new test set, so as to simulate the distortion constraint on watermark. Consequently, we have a total of nine test sets.

- *Metrics*

Eight metrics are compared, including: (1) VIF and MSSIM (the multi-scale version of SSIM), which exhibit an outstanding performance in the previous study [25] [32]; (2) PSNR, WPSNR and CPA, which are either commonly used metrics in watermarking community or devised for watermarking specially; (3) LSE, DCTune and the proposed SOS metrics, which the comparison of watermarking in Sec. IV are based on. The codes of VIF and MSSIM are in the MeTrixMux tool [33], while WPSNR and DCTune are in the CheckMark Benchmarking tool [18]. We choose $a = 5$ and $h = 5$ for SOS as well as $b = 8$ for LSE. This parameter setting makes SOS and LSE achieve the best overall performance, yet a slight adjustment of the parameters will not influence the performance rank since the metric performance is not sensitive to the parameters.

The result of metric performance is presented in Table I, where the best correlations for each test set are marked bold. SOS performs quite well, and it is the only metric which always keep its difference from the highest correlation within 0.03 for all the data sets. The performance of VIF and MSSIM are fairly stable too, yet they do a little worse in the databases for watermarking than those for coding. On the contrary, CPA does well in the databases for watermarking but not for coding. LSE and WPSNR have a similar performance, both of which are good at BA and Meerwald but poor for FourierSB. DCTune sometimes is even inferior to PSNR. SOS obviously outperforms LSE and DCTune. Compared to PSNR, WPSNR and CPA, VIF and MSSIM seem more suitable for the benchmarking of watermarking.

V. PERCEPTUAL WATERMARKING

A. Perceptual Spread Spectrum Watermarking

In this paper we focus on the watermark embedding step of spread spectrum watermarking [34]. Without loss of generality, we embed one bit of message m, $m \in \{-1, +1\}$, into a cover image X. The watermarked image Y is obtained by $Y = X + W$; to generate the watermark pattern W, a pseudo-random carrier P is firstly modulated by m and then is weighted by watermark strength, while the key k controls the generation of P and ensures the security of the watermark. Consequently, W highly correlates with mP. Denote the correlation by

$$c = m\vec{P}^{\mathrm{T}}\vec{W}/(\sigma_P^2 \cdot q) \quad (9),$$

where $\vec{P}, \vec{W} \in \Re^q$; $\sigma^2{}_P$ is the variance of P and often normalized to 1. The correlation c reflects the global watermark strength. The embedding process can be expressed in the vector form:

$$\vec{Y} = \vec{X} + \vec{W} \quad (10).$$

The watermarked image is possibly corrupted by attacking, which is usually modeled as additive noise:

$$\vec{Z} = \vec{Y} + \vec{N} \quad (11).$$

To detect the watermark in a blind way, the watermark decoder uses the key to recover *P*, and then usually makes linear correlation detection (LCD) as (12), without accessing the cover image:

$$\tilde{c} = \vec{P}^{\mathrm{T}}\vec{Z} = \vec{P}^{\mathrm{T}}\vec{W} + \vec{P}^{\mathrm{T}}\vec{X} + \vec{P}^{\mathrm{T}}\vec{N} \quad (12).$$

The last two terms in Eq. (12) can be neglected due to the independency between *X* and *P* and the independency between *N* and *P*. Finally, the watermark message is estimated according to

$$\tilde{m} = \begin{cases} \mathrm{sign}(\tilde{c}) & |\tilde{c}| > \mathrm{threshold} \\ \mathrm{null} & \mathrm{else} \end{cases} \quad (13).$$

To embed multiple bits of message $\{m_1, m_2, \ldots, m_K\}$, the cover image can be divided into *K* nonoverlapping regions $\{\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_K \in \Re^{N_K}\}$, and each region is embedded with one bit of message. That is, the watermark payload can be increased by reducing the repetition rate of one-bit watermark embedding.

In principle, the watermark robustness is evaluated by the average bit error rate (BER) over a set of cover images at the decoder, yet the BER can not be simply compared due to four factors.

- *Watermark attack.* Watermark robustness is a concept closely related to the watermark attack. Different type and strength of attacks may affect the BER to various extents.
- *Watermark payload*. The BER also depends heavily on the watermark payload. For the watermark designer, the payload setting is a compromise due to the application requirements and the channel environments.
- *Detection threshold.* The threshold in (13) influences the BER and the false alarm rate (i.e. reporting watermark from innocent images) in opposite way. To achieve a low BER and avoid being swapped by false alarm, the threshold setting is a compromise for the watermark decoder.
- *Smart detector*. The smart detecting methods, which are often based on the prior knowledge about the embedding methods, may detect watermarks more efficiently and thus improve the BER. This will be discussed in Sec. VII.

To conclude, the robustness is a characteristic dependent on several prerequisites. It is still very challenging for the embedder to predict the BER. We resort to measure the robustness in a simple way. For the embedder who has little knowledge about the potential attacks or the smart detectors, maximizing the correlation of (9) is an intuitive strategy to obtain a robust watermark. Actually, the BER is derived to decrease with the correlation of (9) in assumption that the cover image and the attack noise are independent Gaussian signals [35]. Maximizing correlation is also adopted in [19, 36]. Similarly, we fulfill Problem (1) by defining the *expected robustness* as

$$R(X,Y;m,k) = m\vec{P}^{\mathrm{T}}\vec{W}/(\sigma_P^2 q) \quad (14).$$
$$W = Y - X$$

R is a linear function with *Y*. With the distortion definition of (6), the objective function of Problem (1) is a quadratic with respect to *Y*. The solution to (1) can be derived directly.

At the decoder, the correlation of (12) can roughly capture the *survival robustness* after attacking. The absolute value of correlation usually decreases with the attack strength, and the descending trend reflects the robustness against such type of attack. In this way, we avoid bothering to try all watermark payloads and the detection thresholds.

B. Baseline for comparison

To evaluate watermarking, the characteristics often include invisibility, robustness, capacity, complexity, etc. Capacity is defined as the maximal payload of the watermark to be successfully decoded when the prerequisites including the invisibility constraint, the attack set and the BER constraint are given [37]. It is difficult to compare capacity, because it needs to enumerate and fix the prerequisites above. On the contrary, the respective comparisons of the invisibility and the robustness are helpful in capturing the capacity. All the watermarks to be compared are fast to be embedded and decoded. Due to above reasons, the invisibility and the robustness are the mainly compared. For fair comparison, all the watermarks satisfy a fixed correlation at the embedder:

$$m\vec{P}^{\mathrm{T}}\vec{W}/(\sigma_P^2 q) = \gamma_c \quad (15).$$

At this baseline, the watermarks have equal global strength before attacking, i.e. the same expected robustness. We can compare the distortion by the third-party metrics and compare the survival robustness after attacking according to the absolute correlation of (12).

It is easy to satisfy Constraint (15) for LSE based watermark by simply following the scheme in [15]. For DCTune based watermark, we heuristically make the watermark strength proportional to the $M_{i,j,k}$ in (5) and set the global watermark strength according to Constraint (15). This strategy is often adopted by the watermark based on the HVS mask [1, 17]. For the proposed SOS watermark as well as the LSE and heuristic DCTune watermark above, we always seek the solution to

$$\min_{Y} D(X,Y) \quad \text{s.t.} \quad R(X,Y;m) = \gamma_c \quad (16),$$

where *R* is defined as (14) and *D* is defined as (2) with $\ell = 2$. Problem (16) has the solution:

$$\vec{Y} = \frac{\gamma_c \mathbf{H}\mathbf{M}^{-2}\mathbf{H}^{\mathrm{T}}}{\vec{P}^{\mathrm{T}}\mathbf{H}\mathbf{M}^{-2}\mathbf{H}^{\mathrm{T}}\vec{P}}\vec{P} + \vec{X} \quad (17).$$

In (7), the watermark is a pixel-by-pixel weighted sequence $\vec{P}$, where $\mathbf{H}\mathbf{M}^{-2}\mathbf{H}^{\mathrm{T}}$ is the pixel-wise weights after perceptual shaping, and $\gamma_c / \left(\vec{P}^{\mathrm{T}}\mathbf{H}\mathbf{M}^{-2}\mathbf{H}^{\mathrm{T}}\vec{P}\right)$ is the global weight to satisfy Constraint (15). The only difference among the three kinds of watermarks is the design of **H** and **M**.

VI. EXPERIMENTS FOR WATERMARKING

This section is devoted to comparing the three watermarking schemes based on the LSE metric, DCTune metric, and SOS metric, which will be named respectively as LSE watermark, DCTune watermark and SOS watermark in the rest of this paper. The experiment configuration is described below:

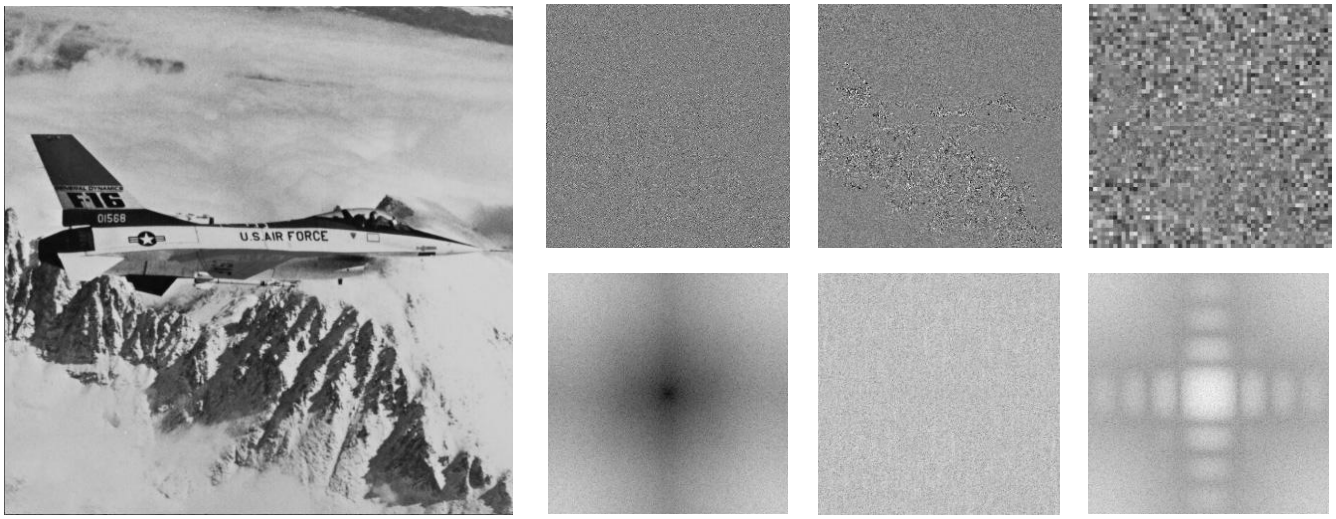


Fig. 1. Original *F16* and the corresponding three types of watermark (Left column: SOS; middle column: LSE; right column: DCTune. Top row: difference in spatial domain between the watermarked image and the cover one; Bottom row: spectra of the error image. The six images are equalized for display purpose).

- *Test images*

29 test images are the full set of the StirMark image database [38], which are collected especially for watermark benchmark covering images with texture and detail, images with lines and edges, smooth bright images, dark images, medical images, computer generated images, etc. The color components of images, if they exist, are removed before use.

- *Watermark strength*

Three levels of γ_c (1, 3, and 5) are used and watermarks are embedded in weak, moderate and strong strength respectively.

- *Attack type*

Gaussian noise, JPEG compression, Wiener filtering and the copy attack [18] are considered. Image enhancement (adjustment of image contrast and mean) makes little impact on spread spectrum watermarks and thus are not discussed. Spread spectrum watermarks are usually fragile to protocol attacks and desynchronization attacks (e.g. cropping and geometrical attack). More sophisticate strategy is necessary to resist those attacks, which are out of the focus in this paper.

A. Distortion

Fig. 1 shows the watermark pattern for image *F16*. In pixel domain, the SOS watermark and the LSE watermark concentrate on the edges and texture regions, while the distribution of the DCTune watermark is relatively even. In spatial frequency domain, the SOS watermark concentrates in high frequency, the DCTune watermark show special distribution in the vertical and horizontal low frequency, while the LSE watermark appears "white" with nearly even spectra.

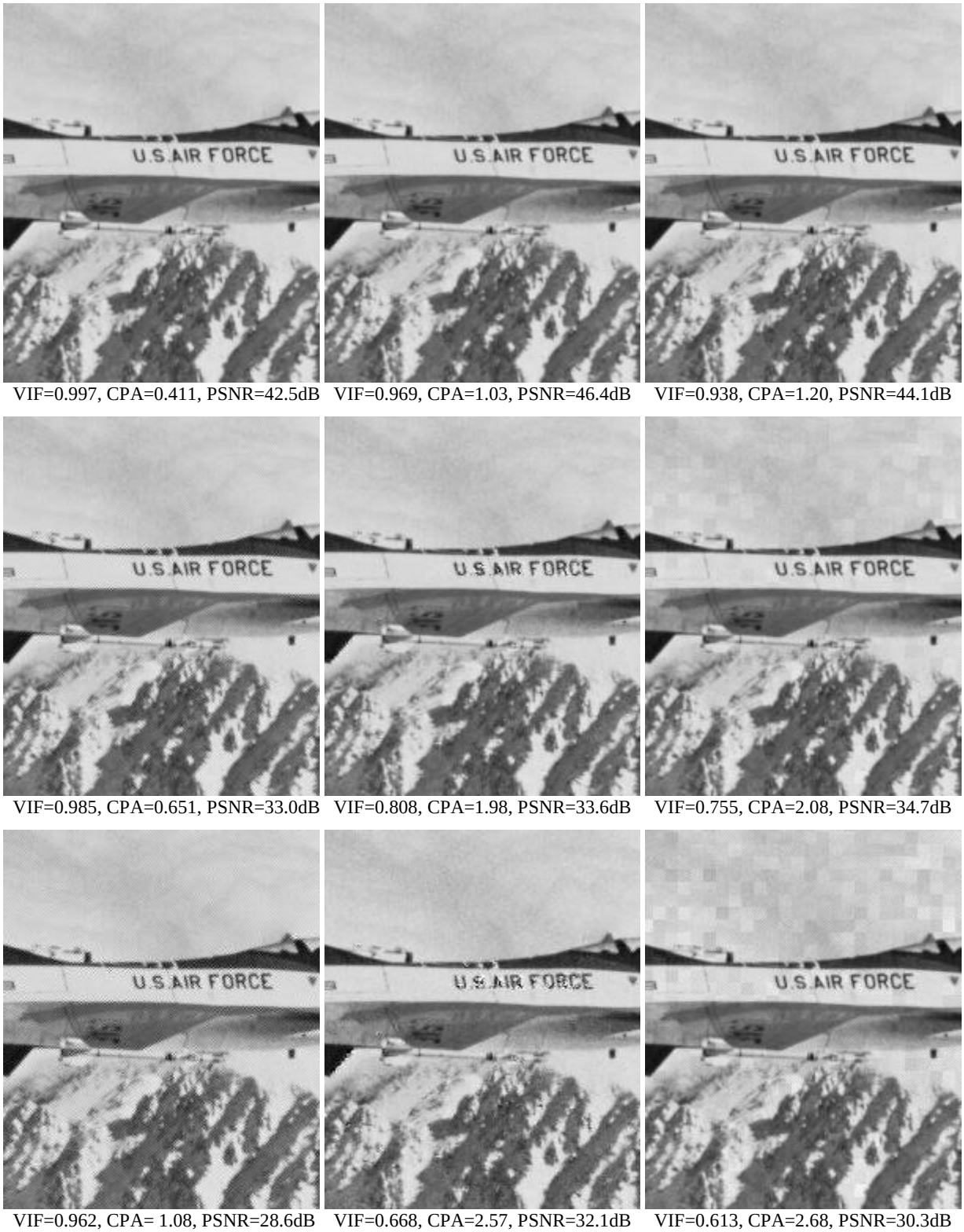
Fig. 2 shows the watermarked results with γ_c of 1, 3 and 5 respectively. For clear comparison, we show the central parts of the results. With scrutiny, we could discriminate the three types of watermarked images from the cover image when γ_c is above 3. However, the LSE watermark impairs sharp edges and texts and the DCTune yields blocking artifacts, which are more annoying than the distortion introduced by the SOS watermark. In Fig. 2, the VIF, CPA and PSNR of each entire watermarked image are also listed as reference to their perceptual distortion and watermark strength respectively. However, the more strict comparison on perceptual distortion is conducted below.

We embedded the three types of watermarks to the 29 StirMark images with three levels of strength, repeated each configuration of random watermark by 10 times, and generated 2610 (=3 ×29 ×3 ×10) watermarked image finally. In Fig. 3, each vertical bar depicts the range of the objective scores which are assigned to the 290 images with the same type and strength of watermark, and the marker depicts the mean of the scores. VIF, MSSIM and WPSNR predict the better quality by the higher scores while CPA on the contrary. Under the evaluation by VIF, MSSIM and CPA, the SOS watermarked images have the best quality overall. The WPSNR metric is quite similar to the LSE metric, so the LSE watermark is destined to keep the most undetectable among the three ones when being measured by WPSNR. With the additional consideration of the comparison in Table I, it is concluded that the SOS watermark is less visible.

B. Correlation (Robustness)

Fig. 4 shows the results of watermark (survival) robustness, which is measured by the correlation of (12). The previous 2610 watermarked images suffer from the four types of attacks respectively. Each vertical bar depicts the range of the correlations which are possessed by the 290 attacked images with the same type and strength of watermark, and the marker depicts the mean of the correlations. The solid line stands for the watermark strength of 1, the dashed line for 3, and the dotted line for 5.

The three types of watermarks are all immune to the Gaussian noise and the copy attack, since the average correlation keeps almost consistent after attack. They are also resilient to JPEG compression to some extent. Specifically, at the high watermark strength (γ_c = 5), the SOS watermark is a little more robust than the other two ones. For the Wiener filtering, the LSE watermark is the winner. The SOS metric appears more fragile to Wiener filtering than others, yet its advantage will be clarified in the next section.



VIF=0.997, CPA=0.411, PSNR=42.5dB VIF=0.969, CPA=1.03, PSNR=46.4dB VIF=0.938, CPA=1.20, PSNR=44.1dB

VIF=0.985, CPA=0.651, PSNR=33.0dB VIF=0.808, CPA=1.98, PSNR=33.6dB VIF=0.755, CPA=2.08, PSNR=34.7dB

VIF=0.962, CPA= 1.08, PSNR=28.6dB VIF=0.668, CPA=2.57, PSNR=32.1dB VIF=0.613, CPA=2.68, PSNR=30.3dB

Fig. 2. Watermarked *F16* (from left to right: SOS, LSE, and DCTune. Top row: $\gamma_c = 1$; middle row: $\gamma_c = 3$; bottom row: $\gamma_c = 5$)

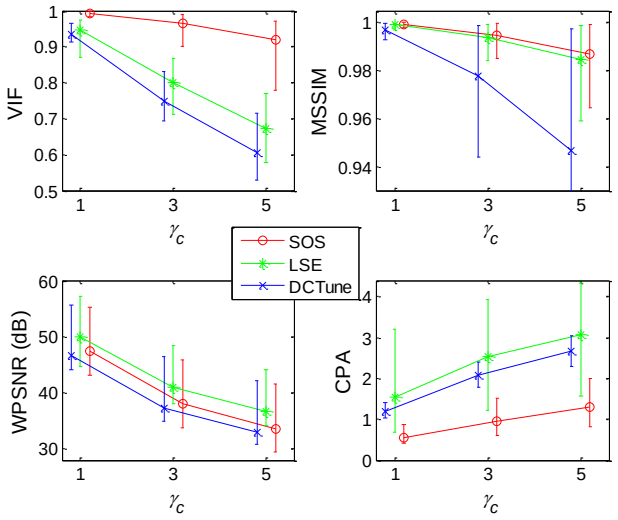


Fig. 3. Objective quality of the StirMark images embedded with three types of watermark. (Better quality corresponds to higher values of VIF, MSSIM and WPSNR and lower CPA scores)

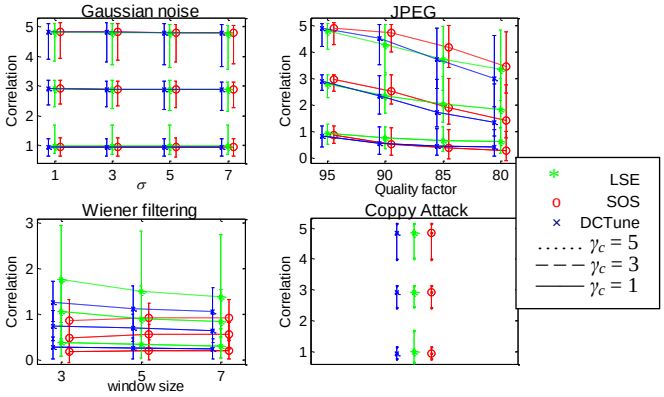


Fig 4. Survival robustness of the three types of watermarks on the SirMark images. (Higher correlation for better robustness)

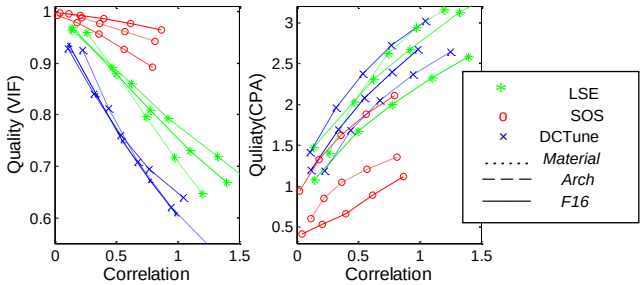


Fig. 5. Curves of the quality of the image embedded with the three types of watermarks versus the correlation after Wiener filtering attack (window size is 3).

C. *Correlation-Quality performance*

Due to the results above, the SOS watermark achieves the least distortion, the comparable robustness against the Gaussian noise, the JPEG compression and the copy attack, as well as a poor robustness against Wiener filtering, given the same correlation at the embedder. Such robustness analysis is not thorough because the watermarks are not compared at the same baseline of invisibility. Aforementioned, the optimal watermark is a tradeoff between the invisibility and the robustness. We suggest using the Correlation-Quality curve to evaluate the comprehensive performance of watermark, just like the Rate-Distortion curve for image coding. To be specific, the attack strength is preassigned, the survival robustness is still measured by Definition (12), while the invisiblity of the watermarked image (before attacking) is evaluated by an accurate third-party metric. For instance, Fig. 5 shows the Correlation-Quality performance of the three types of watermarks on the three StirMark images (*Material*, *Arch* and *F16*), where VIF and CPA is used as the quality metric and Wiener filtering with window size of 3 is assigned. In Fig. 5, the point on the right corresponds to the high correlation and robustness; the higher position in the left figure and the lower position in the right one imply the better quality. The SOS watermarks have the best Correlation-Quality performance therein. In other words, the SOS metric can improve the robustness that is shown in Fig. 4, by taking full advantage of the perceptual distortion constraint. In summary, the SOS watermark promises a better comprehensive performance than the LSE and the DCTune watermarks when exploring the balance between the watermark invisibility and the watermark robustness.

VII. Proposed Detection Scheme

In above comparative study, a simple detection method, LCD, is employed. This is usually efficient in an open industrial society, where watermark decoders have little prior knowledge about watermark embedders and the interoperability between the different spread-spectrum watermarking algorithms can be supported. However, for a closed society, a more sophisticated detector associated with the watermark embedding method may improve the performance. For example, the designers of LSE watermark proposed to use LS filter to estimate the masked watermark before performing the correlation-based detector [15]; Liu developed the locally optimum detector (LOD) to detect JND-based watermarks, where the Bayesian hypothesis testing (BHT) is employed according to the generalized Gaussian distribution (GGD) model on the frequency transform coefficients of natural images [39]. In this section, we present a KLT-based detector associated with the SOS watermark. The detector firstly estimates the SOS watermark by filtering the recieved image and then performs the normalized LCD between the estimated watermark and the watermark carrier.

One interesting feature of the SOS metric is that it shapes the SOS watermark pattern into distinct KLT spectra from the cover image. As aforementioned, the SOS watermark pattern exhibits the eigen-values being reciprocal to the cover image's eigen-values, that is, the SOS watermark pattern concentrates most of its energy in the least principal components of the cover image. As illustrated in Fig. 6, the samples of original *F16* (the curve marked by circles) have descending eigen-values from the most principal KLT basis to the least one. The samples of the *F16* with a SOS watermark (the curve marked by squares) exhibit approximately descending eigen-values except having a bigger tail. This is because it has

been added with a SOS watermark which has ascending eigen-values[1] (the curve marked by cross). Therefore, it is reasonable to estimate SOS watermark pattern by "high-pass" filtering the received image in the KLT domain of the cover image. Although the KLT bases of the cover image are not available for a blind watermark detector, it is found that the SOS watermarked images have approximately equal KLT bases with those of the cover images, except that the least principal KLT bases are regrouped in order. Consequently, it is feasible to do filtering in the KLT domain of the received image. The filter shows a high-pass frequency response function as the curve marked by dots in Fig. 6. Note that the filter is defined in the KLT domain of the received image; we approximately show it in that of the cover image for illustration purpose.

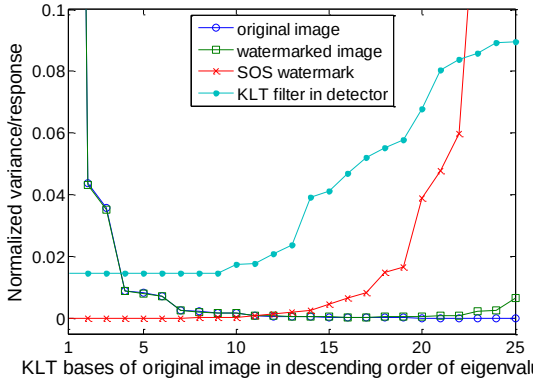


Fig. 6. Demonstration of the signal spectra and the filter response in the KLT domain of StirMark image *F16*.

By the eigendecomposition of the covariance matrix of the received image samples, the decoder has $\mathbf{S}' = \mathbf{U}'\,\mathbf{\Lambda}'\,\mathbf{U}'^{\mathrm{T}}$. The columns of $\mathbf{U}'^{\mathrm{T}}$ represent the KLT bases of the received image and $\mathbf{\Lambda}'$ has a^2 eigenvalues, $\lambda'_1\,\lambda'_2\,\ldots\,\lambda'_{a^2}$, in a descending order as its diagonal elements. Filtering the received image and estimating the SOS watermark is made by

$$\tilde{\tilde{w}}_k = \mathbf{U}'\mathbf{F}\mathbf{U}'^{\mathrm{T}}\vec{z}_k \qquad (18).$$

The diagonal matrix $\mathbf{F}$ captures the frequency response function, having the first η ($\eta < a^2$) equal diagonal elements, $1/\lambda'_\eta$, and the last ($a^2 - \eta$) diagonal elements, $1/\lambda'_{\eta+1}\ 1/\lambda'_{\eta+2}\ \ldots\ 1/\lambda'_h$, in an ascending order. The diagonal elements are illustrated by the curve marked by dots in Fig. 6. $\vec{z}_k$ is the vectorized received image block z_k and $\tilde{\tilde{w}}_k$ is the estimation of the watermark pattern block in the vector form. We set $\eta = 9$ empirically, while $a = 5$ as same as the watermark embedder. Then, the normalized correlation between the estimated watermark pattern and the watermark carrier is calculated to determine the watermark message:

$$\tilde{c}_k = \left(p_k{}^{\mathrm{T}}\tilde{\tilde{w}}_k\right) \Big/ \left(\sigma_P \cdot \sigma_{\tilde{W}}\right) \qquad (19).$$

σ_P and $\sigma_{\tilde{W}}$ is the standard deviation of the watermark carrier P and the watermark pattern $\tilde{W}$, which contain one bit of the watermark message. Finally, the watermark message is estimated according to

$$\tilde{m} = \begin{cases} \mathrm{sign}\left(\sum_k \tilde{c}_k\right) & \left|\sum_k \tilde{c}_k\right| > \mathrm{threshold} \\ \mathrm{null} & \mathrm{else} \end{cases} \qquad (20).$$

The detection performance of the KLT-based detector and the normalized LCD are compared on the StirMark images without attacks. Embedding one bit message "+1" into every 40×40 nonoverlapping block of an image, we collected a total of 7527 image blocks from 29 original StirMark images as well as 29 watermarked images, and then input them to the two detectors respectively. The distributions of outputs are shown in Fig. 7. A significant improvement is observed by using KLT-based detector.

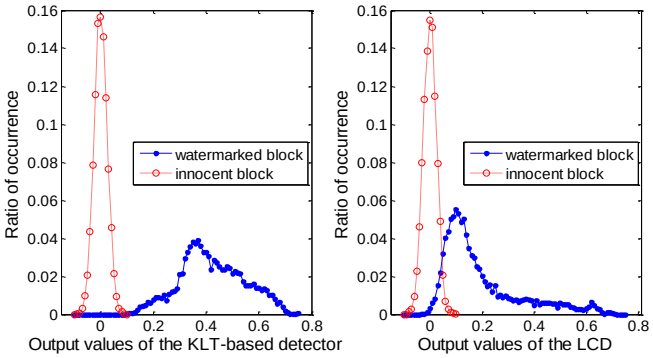


Fig. 7. Output distributions of the detectors for the 7527 40×40 StirMark image blocks with SOS watermarks.

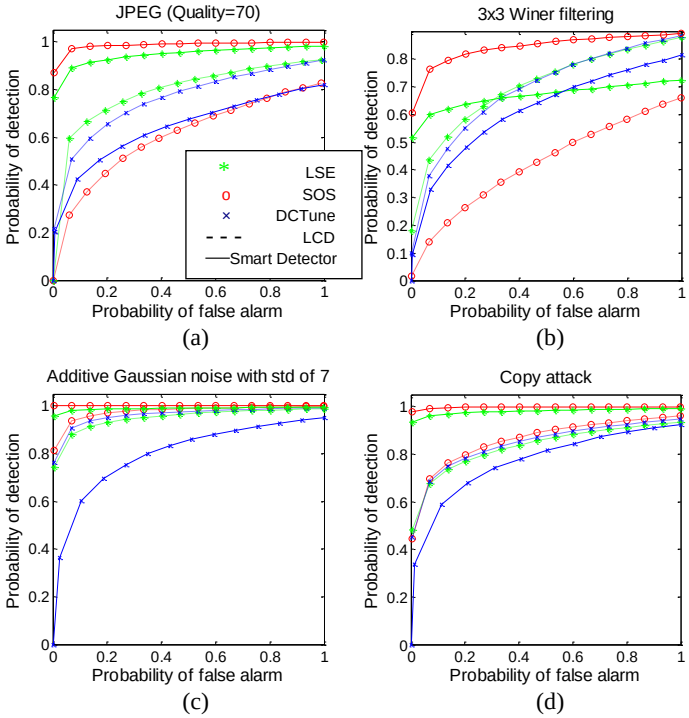


Fig. 8. ROC curves for the 7527 40×40 StirMark image blocks after (a) JPEG compression with the quality factor being 70, (b) Wiener filtering with window size being 3, (c) additive Gaussian noise with std of 7, and (d) copy attack.

[1] Since the watermark and the watermarked image do not have exactly the same KLT bases as the cover image, their eigen-value is calculated as the variance of their sample projections along the corresponding KLT basis.

In Sec. VI-B and VI-C, we use the correlation of (12) to evaluate the robustness of the watermarks in general. In this section, we further compare the robustness by the ROC curve at the decoder for the particular watermark attacks and certain watermark payloads. Both the simple LCD and the smart detectors are considered among the SOS, the LSE and the DCTune watermarks. For the smart detections, the KLT-based detector is integrated for the SOS watermark, the LS-filter-based detector for the LSE [15], and the LOD for the DCTune [39]. For fair comparison, we calculate the LOD if among the image region more than 20% DCT coefficients have the magnitudes above the Watson's mask (see $M_{i,j,k}$ in Eq. (5); using Watson's mask as the threshold is proposed in [39]), otherwise still calculate the LCD because the LOD does not work well for small coefficients as pointed out in [39]. We adjust the detecting threshold, which makes the LCD and the LOD achieve the equal false alarm rates, and then average their probabilities of detection as the final one.

The baseline of a moderate watermark strength (γ_c = 3) is kept at the embedders for the three types of watermarks. Again, the original StirMark images and their watermarked versions with a message "+1" in every 40×40 block are input to the detectors after being attacked. JPEG compression with a quality factor being 70, Wiener filtering with a window size being 3, additive Gaussian noise with the standard deviation of 7 and the copy attack are considered respectively. For each case of the attacks, the ROC curves of the three watermarks are plotted in Fig. 8, where the dashed curves stand for employing the simple LCD and the solid curve for the corresponding smart detector. Two main observations are found from the results:

- With the LCD detection, the SOS watermark can be less efficiently detected than the other two schemes for JPEG compression and Wiener filtering.
- With the smart detection, the SOS watermark obtains the most performance gain and outperforms the others for all the four attacks.

The significant incremental performance of the KLT-based detector confirms that the distinct KLT spectra of the SOS watermark can facilitate the watermark detection. Given the same watermark strength at the embedder, the SOS watermark has kept the least visibility in Sec. VI-C, and has also exhibited the most detection efficiency in this section. Therefore, taking advantage of the KLT based detector, the SOS watermark achieves the best comprehensive performance against the attack sets having been considered. It should be cautiously pointed out that the SOS watermark is vulnerable to a "low-pass" filtering in the KLT domain, just like every perceptual watermarking scheme suffering from a particular attack associated with the corresponding visual mask in the corresponding transform domain.

VIII. Conclusion

For designing a robust watermark, it is very important to exploit the perceptual distortion in line with human perception. In this paper, we propose the SOS metric, which provides two advantages:

- *Accurate.* It predicts image quality in a way that highly correlates with the subjectively-rated databases, and guides watermarking by introducing the distortion as invisible as possible.
- *Simple.* It ensures a closed-form solution to the optimal watermark.

This paper also suggests using the correlation-quality curve to objectively evaluate the performance of robust watermarks, where the quality of watermarked images should be rated by third-party metrics. With the explicit SOS metric, the proposed watermark achieves a good correlation-quality performance and can make a satisfactory tradeoff between the robustness and the distortion, compared to two typical watermarking schemes. Moreover, the SOS watermark can be efficiently detected by exploiting its distinct KLT spectra from those of the cover image.

Although this paper only discusses the spread spectrum watermark, the SOS metric is promised to have wider applications like:

- *QIM watermark.* QIM (Quantization Index Modulation) [40] can be also formulated as problem (1) with an adequate definition of R, and thus employ the SOS metric.
- *Image Matching*. PCA (Principal Component Analysis) method is frequently used for image matching. The SOS metric is based on PCA too. Its accuracy on subjectively-rated databases is good. It may be applied to texture synthesis, image inpainting, etc.

Acknowledgement

Dr. Wenyu Liu is the corresponding author of this paper. The authors thank Dr. F. Autrusseau for kindly sharing the source code of the CPA metric, as well as Dr. Wei Liu and Prof. Wenjun Zeng for the implementation of the LOD.

References

[1] C. I. Podilchunk, and W. Zeng, "Digital image watermarking using visual models," in *Human Vision and Electronic Imaging II*, 1997, pp. 100-111.

[2] M. Kutter, and S. Winkler, "A vision-based masking model for spread spectrum image watermarking," *IEEE Trans Image Process,* vol. 11, no. 1, pp. 16 - 25, 2002.

[3] H. Qi, D. Zheng, and J. Zhao, "Human visual system based adaptive digital image watermarking," *Signal Processing,* vol. 88, no. 1, pp. 174 - 188, 2008.

[4] W. S. Lin, "Computational Models for Just-Noticeable Difference," *Digital video image quality and perceptual coding*, H. R. Wu and K. R. Rao, eds., Boca Raton: CRC Press, 2005.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh *et al.*, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing,* vol. 13, no. 4, pp. 600-612, 2004.

[6] S. Daly, "The Visible Differences Predictor - an Algorithm for the Assessment of Image Fidelity," *Human Vision, Visual Processing, and Digital Display III*, vol. 1666, pp. 2-15, 598, 1992.

[7] A. B. Watson, "DCTune: A technique for visual optimization of DCT quantization matrices for individual images," *Society for Information Display Digest of Technical Papers, XXIV, 946-949*, 1993.

[8] A. B. Watson, G. Y. Yang, J. A. Solomon *et al.*, "Visibility of wavelet quantization noise," *IEEE Trans. Image Processing,* vol. 6, no. 8, pp. 1164-1175, Aug, 1997.

[9] U. Grenander, and G. Szego, *Toeplitz forms and their applications*, 2nd ed. ed., New York: Chelsea, 1981.

[10] A. B. Watson, and L. Kreslake, "Measurement of visual impairment scales for digital video," in *Human Vision, Visual Processing, and Digital Display, Proc. SPIE*, 2001.

[11] H. R. Sheikh, and A. C. Bovik, "Image information and visual quality," *IEEE Trans Image Processing*, vol. 15, no. 2, pp. 430-444, 2006.
[12] Z. Wang, Q. Li, and X. L. Shang, "Perceptual image coding based on a maximum of minimal structural similarity criterion," *International Conference on Image Processing*, vol. 1-7, pp. 685-688, 2007.
[13] S. Voloshynovskiy, S. Pererira, A. Herrigel *et al.*, "Generalized watermarking attack based on watermark estimation and perceptual remodulation," in *SPIE Security and Watermarking of Multimedia Contents*, 2000, pp. 358 - 370.
[14] M. Carosi, V. Pankajakshan, and F. Autrusseau, "Towards a simplified perceptual quality metric for watermarking applications," in *ICIP*, 2010.
[15] I. G. Karybali, and K. Berberidis, "Efficient spatial image watermarking via new perceptual masking and blind detection schemes," *IEEE Trans. Inf. Forensices Security*, vol. 1, no. 2, pp. 256 - 274, 2006.
[16] S. Voloshynovskiy, A. Herrigel, N. Baumgaertner *et al.*, "A stochastic approach to content adaptive digital image watermarking," *Information Hiding, Proceedings*, vol. 1768, pp. 211-236, 2000.
[17] Q. Li, and I. J. Cox, "Using perceptual models to improve fidelity and provide resistance to valumetric scaling for quantization index modulation watermarking," *IEEE Trans. Inf. Forensices Security*, vol. 2, no. 2, pp. 127 -139, 2007.
[18] S. Pereira, S. Voloshynovskiy, M. Madueno *et al.*, "Second generation benchmarking and application oriented evaluation," in *Information Hiding Workshop*, Pittsburgh, PA, USA, LNCS, vol. 2137, pp. 340 - 353, 2001.
[19] H. Oktay Altun, A. Orsdemir, G. Sharma *et al.*, "Optimal spread spectrum watermark embedding via a multistep feasibility evaluation," *IEEE Trans. Image Processing*, vol. 18, no. 2, pp. 371 - 387, 2009.
[20] B. Julesz, E. N. Gilbert, L. A. Shepp *et al.*, "Inability of humans to discriminate between visual textures that agree in second-order statistics - revisited," *Perception*, vol. 2, pp. 391 - 405, 1973.
[21] A. Hyvärinen, G. Borgefors, J. Hurri *et al.*, "Natural Image Statistics A Probabilistic Approach to Early Computational Vision," *Computational Imaging and Vision, 39.*, Springer London, pp. 106 - 107, 2009.
[22] P. C. Mahalanobis, "On the generalised distance in statistics," in *Proc. of the National Institue of Sciences of India*, 1936, pp. 49 - 55.
[23] P. Le Callet, F. Autrusseau, and P. Campisi, "Visibility control and quality assessment of watermarking and data hiding algorithms," *Multimedia Forensics and Security*, C. T. LI, ed., pp. 163 - 192: IGI Global snippet, 2008.
[24] ITU-R-Recommendation-BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," 2002.
[25] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3440-3451, 2006.
[26] H. R. Sheikh, "LIVE Image Quality Assessment Database, Release 2" [online]. Available: http://live.ece.utexas.edu/research/quality, 2005.
[27] N. Ponomarenko, "TAMPERE IMAGE DATABASE 2008 version 1.0" [online]. Available: http://www.ponomarenko.info/tid2008.htm, 2008.
[28] D. M. Chandler, "CSIQ image database" [online] Available: http://vision. okstate.edu/index.php?loc=csiq, 2010.
[29] F. Autrusseau, P. Le Callet, "FourierSB database" [online] Available: http://www.irccyn.ec-nantes.fr/~autrusse/Databases/FourierSB/
[30] F. Autrusseau, P. Le Callet, "BA database" [online] Available: http://www.irccyn.ec-nantes.fr/~autrusse/Databases/BA/
[31] F. Autrusseau, P. Le Callet, "Meerwald database" [online] Available: http://www.irccyn.ec-nantes.fr/~autrusse/Databases/Meerwald/
[32] N. Ponomarenko, V. Lukin, A. Zelensky *et al.*, "TID 2008 - a database for evaluation of full reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, pp. 30 - 45, 2009.
[33] M. Gaubatz, "MeTrix Mux visual quality assessment package", [online] Available: http://foulard.ece.cornell.edu/gaubatz/metric_mux/
[34] I. J. Cox, J. Killian, F. T. Leighton *et al.*, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1673 - 1687, 1997.
[35] H. S. Malvar, and D. A. F. Florencio, "Improved spread spectrum: a new modulation technique for robust watermarking," *IEEE Trans. Signal Processing*, vol. 51, pp. 898 - 905, 2003.
[36] F. Pereira, S. Voloshynovskiy, and T. Pun, "Optimal transform domain watermark embedding via linear programming," *Signal Processing*, vol. 81, no. 6, pp. 1251 - 1260, 2001.
[37] P. Moulin, and A. Ivanovic, "The zero-rate spread-spectrum watermarking game," *IEEE Trans. Signal Processing*, vol. 51, no. 4, pp. 1098 - 1117, 2003.
[38] "StirMark image database 2.0" [online] Available: http://www.petitcolas.net/fabien/watermarking/image_database/index.html
[39] W. Liu, L. Dong, and W. Zeng, "Optimum detection for spread-spectrum watermarking that employs self-masking," *IEEE Trans. Inf. Forensices Security*, vol. 2, no. 4, pp. 645 - 654, 2007.
[40] B. Chen, and G. W. Wornell, "Quantization idex modulation methods: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423 - 1443, 2001.

Fan Zhang (M'11) received his B. E. and Ph. D. degrees from Huazhong University of Science and Technology, China, in 2002 and 2008 respectively, both in electronic and information engineering. He was a visiting student at the Nanyang Technological University in 2008 and a postdoctor at the Chinese University of Hong Kong from 2009 to 2010. He is a research engineer in Technicolor Research (Beijing) since 2010. He organized the special session in Asia Pacific Signal and Information Processing Association (APSIPA 2011). His research interests are quality of experience and perceptual watermarking.

Wenyu Liu (M08) received the B.S. degree in computer science from Tsinghua University, Beijing, China, in 1986, and the M.S. and Ph.D. degrees in electronics and information engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 1991 and 2001, respectively. He is now a Professor and Associate Dean of the Department of Electronics and Information Engineering, HUST. His current research areas include multimedia information processing, and computer vision.

Weisi Lin (M'92–SM'98) received the B.Sc. degree in electronics and the M.Sc. degree in digital signal processing from Zhongshan University, Guangzhou, China, in 1982 and 1985, respectively, and the Ph.D. degree in computer vision from King's College, London University, London, U.K., in 1992. He taught and conducted research at Zhongshan University, Shantou University (China), Bath University (U.K.), the National University of Singapore, the Institute of Microelectronics (Singapore), and the Institute for Infocomm Research (Singapore). He has been the Project Leader of 13 major successfully-delivered projects in digital multimedia technology development. He also served as the Lab Head, Visual Processing, and the Acting Department Manager, Media Processing, for the Institute for Infocomm Research. Currently, he is an Associate Professor in the School of Computer Engineering, Nanyang Technological University, Singapore. His areas of expertise include image processing, perceptual modeling, video compression, multimedia communication and computer vision. He holds ten patents, edited one book, authored one book and five book chapters, and has published over 160 refereed papers in international journals and conferences. He believes that good theory is practical, so has kept a balance of academic research and industrial deployment throughout his working life.

Dr. Lin is a fellow of Institution of Engineering Technology. He is also a Chartered Engineer (U.K.). He organized special sessions in IEEE International Conference on Multimedia and Expo (ICME 2006), IEEE International Workshop on Multimedia Analysis and Processing (2007), IEEE International Symposium on Circuits and Systems (ISCAS 2010), Pacific-Rim Conference on Multimedia (PCM 2009), SPIE Visual Communications and Image Processing (VCIP 2010), Asia Pacific Signal and Information Processing Association (APSIPA 2011), and MobiMedia 2011.

He gave invited/keynote/ panelist talks in International Workshop on Video Processing and Quality Metrics (2006), IEEE International Conference on Computer Communications and Networks (2007), SPIE VCIP 2010, and IEEE Multimedia Communication Technical Committee (MMTC) Interest Group of Quality of Experience for Multimedia Communications (2011), and tutorials in PCM 2007, PCM 2009, IEEE ISCAS 2008, IEEE ICME 2009, APSIPA 2010, and IEEE International Conference on Image Processing (2010). He is currently on the editorial boards of IEEE Trans. on Multimedia, IEEE SIGNAL PROCESSING LETTERS and Journal of Visual Communication and Image Representation, and four IEEE Technical Committees. He cochairs the IEEE MMTC Special Interest Group on Quality of Experience. He has been on Technical Program Committees and/or Organizing Committees of a number of international conferences.

King N. Ngan received the Ph.D. degree in Electrical Engineering from the Loughborough University in U.K. He is currently a chair professor at the Department of Electronic Engineering, Chinese University of Hong Kong. He was previously a full professor at the Nanyang Technological University, Singapore, and the University of Western Australia, Australia. He holds honorary and visiting professorships of numerous universities in China, Australia and South East Asia.

Prof. Ngan is an associate editor of the Journal on Visual Communications and Image Representation, as well as an area editor of EURASIP Journal of Signal Processing: Image Communication, and served as an associate editor of IEEE Transactions on Circuits and Systems for Video Technology and Journal of Applied Signal Processing. He chaired a number of prestigious international conferences on video signal processing and communications, and served on the advisory and technical committees of numerous professional organizations. Prof. Ngan is a general co-chair of the IEEE International Conference on Image Processing (ICIP) held in Hong Kong in September 2010. He has published extensively including 3 authored books, 5 edited volumes, over 300 refereed technical papers, and edited 9 special issues in journals. In addition, he holds 10 patents in the areas of image/video coding and communications.

Prof. Ngan is a Fellow of IEEE (U.S.A.), IET (U.K.), and IEAust (Australia), and an IEEE Distinguished Lecturer in 2006-2007.