International Journal of Innovative Computing, Information and Control
Volume **9**, Number **7**, July **2013**
ICIC International ©2013 ISSN 1349-4198
pp. 2789–2800

AUDIO WATERMARKING FRAMEWORK USING MULTI-OBJECTIVE PARTICLE SWARM OPTIMIZATION

Hong Peng[1], Zulin Zhang[2], Jun Wang[3] and Peng Shi[4,5]

[1]Center for Radio Administration and Technology Development
[3]School of Electrical and Information Engineering
Xihua University
Chengdu 610039, P. R. China
ph.xhu@hotmail.com

[2]Department of Computer Science
Sichuan University for Nationalities
Kangding 626001, P. R. China

[4]School of Engineering and Science
Victoria University
Melbourne, Vic 8001, Australia

[5]School of Electrical and Electronic Engineering
The University of Adelaide
Adelaide, SA 5005, Australia

Received April 2012; revised August 2012

Abstract. *Aiming at the multi-objective essence of optimal audio watermarking problem, we propose a novel audio watermarking framework in this paper, which can optimally balance all conflicting objectives of the problem, fidelity and robustness against different attacks. In the proposed framework, a multi-objective particle swarm optimization technique based on fitness sharing is applied to search optimal watermarking parameters and Pareto-optimal solutions are used to express the optimal parameters found. In addition, the proposed framework has the following advantages: (i) it can avoid the difficulty of determining optimal weighted factors in the existing single-objective watermarking schemes; (ii) Pareto-optimal solutions can offer the flexibility to select optimal parameters for satisfying different application demands.*
Keywords: Optimal audio watermarking, Audio watermarking framework, Multi-objective optimization, Particle swarm optimization

1. **Introduction.** Audio watermarking is a promising technique that deals with copyright protection of digital audio [1, 2, 3]. Fidelity and robustness are two known performance measures in audio watermarking. However, the two measures are conflicting with each other [4, 5, 6]. How optimally to balance the fidelity and robustness against different attacks? This raises an interesting problem in digital watermarking area, called the optimal watermarking problem.

Existing works have shown that an effective way is to consider the optimal watermarking problem as an optimization problem. Therefore, some known optimization techniques can be used to discuss this problem, such as genetic algorithm (GA) and particle swarm optimization (PSO). Most of the existing works focus on optimal image watermarking, and a few works discuss optimal audio watermarking. In summary, use of the GA/PSO in digital watermarking usually has two purposes: one is to determine optimal watermarking parameters for a watermarking algorithm automatically, such as embedding strengths and thresholds; the other is to search most suitable embedding positions for watermark embedding, such as transform domain coefficients or frequency bands. Huang and Wu [7] proposed an image watermarking scheme in which GA was used to search most suitable embedding positions from DCT coefficient blocks. Another optimal watermarking scheme [8] used the GA to search most suitable frequency bands for watermark embedding. In order to determine optimal watermarking parameters automatically, Kumsawat et al. [9] presented a GA-based watermarking scheme in multiwavelet domain. In addition, Huang and Wu [10] proposed a GA-based framework for watermarking performance enhancement, while Meng et al. [11] presented an adaptive watermarking scheme that incorporated support vector machine (SVM) and GA. Similarly, GA was used in audio watermarking [12]. Besides, optimal watermarking scheme that used the PSO has been developed [13].

In the following, let us analyze the existing watermarking schemes based on GA or PSO (for convenience, called single-objective watermarking schemes below), which convert the optimal watermarking problem to a single-objective optimization problem. As we know, fidelity and robustness are conflicting with each other. Therefore, from the view of objective optimization, the existing single-objective watermarking schemes failed to balance the conflicting objectives optimally. In addition, a weakness exists in the single-objective watermarking schemes: the difficulty of determining optimal weighted factors. Furthermore, we observe from our experiments that different attacks have different effects to performance of a watermarking algorithm and robustness against different attacks are also conflicting with each other in fact. The observation indicates that robustness against each attack should be a performance objective of watermarking algorithm except the fidelity. Thus, the optimal watermarking problem should be essentially a multi-objective optimization problem. The main motivation behind our study is to develop an effective method to solve the optimal audio watermarking problem according to its multi-objective essence.

For this purpose, we propose an audio watermarking framework based on multi-objective particle swarm optimization in this paper. Pareto-optimal concept will be used to characterize the solutions of the optimal audio watermarking problem and a multi-objective particle swarm optimization technique based on fitness sharing will be applied in the proposed audio watermarking framework. The proposed framework can automatically determine optimal watermarking parameters such that the conflicting objectives are optimally balanced. Thus, the framework can avoid the difficulty of determining optimal weighted factor in single-objective watermarking scheme. Pareto optimal set generated by the proposed framework can provide the flexibility to select a most suitable watermarking parameters for practical application demand. In addition, the proposed audio watermarking framework has simple implementation and low computational cost.

This paper is organized as follows. Particle swarm optimization and a multi-objective particle swarm optimization based on fitness sharing are introduced in Section 2. Section 3 describes the proposed audio watermarking framework. In Section 4, experimental results and discussions are given. The conclusions of our work can be found in Section 5.

2. Particle Swarm Optimization and Multi-objective Optimization.

2.1. Particle swarm optimization.

Particle swarm optimization (PSO) is a known population-based optimization technique, first proposed by Kennedy and Eberhart [14, 15]. It was inspired from the social behaviors of bird flocking and fish schooling in search of food. In PSO, potential solutions are expressed by a swarm of particles, and each particle P_m is associated with two vectors: velocity vector $V_m = (v_m^1, v_m^2, \ldots, v_m^D)$ and position vector $X_m = (x_m^1, x_m^2, \ldots, x_m^D)$, where D stands for the dimension of the solution space. Suppose the population's size is M.

PSO algorithm can be summarized as follows. PSO firstly starts with a set of M initial particles, where the velocity and position vectors of each particle are initialized randomly within the corresponding ranges. And then, the following velocity-position model is used to update each particle:

$$v_m^d = w \cdot v_m^d + c_1 r_1 \cdot (pbest_m^d - x_m^d) + c_2 r_2 \cdot (gbest^d - x_m^d), \tag{1}$$

$$x_m^d = x_m^d + v_m^d, \quad (m = 1, 2, \ldots, M, \ d = 1, 2, \ldots, D) \tag{2}$$

where w is inertia weight, c_1 and c_2 are learning factors, r_1 and r_2 are random numbers in $[0, 1]$. In the velocity-position model, $pbest_m^d$ is the best position found by the mth particle so far, while $gbest^d$ is the best position found by all particles so far. During evolution, the objective values (fitness) of the particles are evaluated. The algorithm iterates repeatedly until the satisfaction of some terminating criteria which is problem-dependent.

2.2. **Multi-objective particle swarm optimization.** In this section, we review an important multi-objective particle swarm optimization technique based on fitness sharing, called MOPSO-fs [16, 17], which will be used in our audio watermarking framework. MOPSO-fs utilizes not only the PSO technique to guide the search but also the fitness sharing concept to spread the solution along the Pareto front. The structure of MOPSO-fs algorithm is described as follows.

1) In the first step, the algorithm initializes with random population from an uniform distribution. The external repository is filled with all the nondominated particles. The fitness sharing is calculated for each particle in the repository. According to fitness sharing principle, particles that have more particles in their vicinity are less fit than those that have fewer particles surrounding their vicinity. The fitness sharing for mth particle is given by $f_{sh}^m = 10/nCount_m$, and $nCount_m = \sum_{j=1}^{R} sharing_m^j$. Here, R is the number of particles in the repository and $sharing_m^j$ is derived by

$$sharing_m^j = \begin{cases} 1 - (d_m^j/\sigma_{share})^2, & \text{if } d_m^j < \sigma_{share}, \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

where d_m^j is the Euclidean distance between the mth and the jth particle, while σ_{share} is the radius of the vicinity area of a particle. A high value of f_{sh}^m suggests that the vicinity of mth particle is not highly populated.

2) Provided that a fitness sharing is assigned for each particle in the repository, some particles from the repository are chosen as leaders. These particles are going to be followed by the rest of the particles in the next iteration. The leaders are chosen according to roulette wheel selection, i.e., particles with higher levels of fitness are likely to be selected. This will allow them to explore places less explored in the search space. We randomly select a particle from these leaders, and its position is regarded as $gbest$. The velocity and position for each particle are updated by (1) and (2) respectively.

3) The repository is updated with the current solutions found by the particles. Two criteria are used: dominance and fitness sharing. The particles that dominate those inside the repository will be inserted whereas all solutions dominated will be deleted. Thus, we maintain the repository as the Pareto front found so far. In the case where the repository is full of nondominated particles and a particle nondominated by any in the repository is found, their fitness sharing is compared. If it is better than the worst fitness sharing in the repository, then the new particle replaces the one

with the worst fitness sharing. The fitness sharing of all particles is updated when a particle is inserted in the repository or deleted from the repository.

4) Finally, the memory of each particle is updated according to the criteria of dominance. Therefore, if current particle position dominates previous one, current position replaces previous one in the particle's memory.

3. The Proposed Audio Watermarking Framework.

3.1. The optimized objectives.
In the existing watermarking schemes, fidelity is commonly measured by peak signal-to-noise ratio (PSNR), and normalized cross-correlation (NC) is used to measure the robustness against different attacks. The PSNR and NC are defined as follows.

$$\mathrm{PSNR} = 10\log_{10}\left[\frac{N_1 \times \max_{1\le i\le N_1}|a(i)|^2}{\sum_{i=1}^{N_1}(a(i)-a'(i))^2}\right], \quad \mathrm{NC} = \sum_{i=1}^{N_2}\frac{[w(i)\times w'(i)]}{[w(i)]^2} \tag{4}$$

where $a(i)$ is sample value of original audio signal, $a'(i)$ is sample value of the watermarked audio signal, N_1 is length of audio signal, $w(i)$ is original watermark bit, $w'(i)$ is the extracted watermark bit, and N_2 is length of watermark signal.

As described above, not only fidelity and robustness objectives are conflicting with each other, but also robustness objectives against different attacks are conflicting with each other. Therefore, optimal watermarking problem should be essentially a multi-objective optimization problem. We will apply MOPSO-fs algorithm to solve the multi-objective optimal watermarking problem, where fidelity and robustness against different attacks are its optimized objectives.

Formally, suppose watermarking algorithm considered has D watermarking parameters, $x_1, x_2, \ldots, x_D$, where $\beta_i \le x_i \le \gamma_i$. Moreover, $K-1$ attack methods are considered. Thus, the multi-objective optimal watermarking problem has K objectives, where first objective is the fidelity and other $K-1$ objectives are the robustness against $K-1$ attacks respectively. The multi-objective optimal watermarking problem can be formally expressed as follows.

$$\begin{cases} \max\ (f_1, f_2, \ldots, f_K) \\ f_1 = f_1(x_1, x_2, \ldots, x_D) = \mathrm{PSNR}(x_1, x_2, \ldots, x_D) \\ f_2 = f_2(x_1, x_2, \ldots, x_D) = \mathrm{NC}_1(x_1, x_2, \ldots, x_D) \\ \cdots\cdots \\ f_K = f_K(x_1, x_2, \ldots, x_D) = \mathrm{NC}_{K-1}(x_1, x_2, \ldots, x_D) \end{cases} \tag{5}$$

3.2. Audio watermarking algorithm.
The audio watermarking algorithm uses mean quantization technique, whose idea is similar to Chen and Lin [18]. Suppose $A = \{a(i), 0 \le i < N_1\}$ is an original audio signal, where N_1 is the number of audio samples. The audio signal A is divided into audio frames with length L, $A = \{A(k), 0 \le k < N_1/L\}$, where $A(k)$ is kth audio frame. And then, one level DWT decomposition is accomplished on every audio frame. Low-frequency coefficients of kth audio frame are denoted formally by $C_k = \{c_{ki} \mid 0 \le i < L/2\}$, and their mean value is calculated by $\bar{c}_k = \frac{1}{L/2}\sum_{i=1}^{L/2} c_{ki}$. Let Δ_k be the quantization step. The rule of mean quantization is as follows:

$$\bar{c}'_k = \begin{cases} \lfloor \bar{c}_k/\Delta_k \rfloor \cdot \Delta_k + 3\Delta_k/4 & \text{if } w_k = 1 \\ \lfloor \bar{c}_k/\Delta_k \rfloor \cdot \Delta_k + \Delta_k/4 & \text{if } w_k = 0 \end{cases} \tag{6}$$

where $\bar{c}'_k$ is mean value after quantizing. Thus, low-frequency coefficients of kth audio frame are modulated by

$$c'_{ki} = c_{ki} + (\bar{c}'_k - \bar{c}_k), \quad 0 \le i < L/2, \tag{7}$$

where c'_{ki} $(0 \leq i < L/2)$ are low-frequency coefficients after quantizing.

Let $A' = \{a'(i), 0 \leq i < N_1\}$ be a watermarked audio signal to be tested. A' is divided into audio frames and then DWT transform is accomplished on these audio frames respectively. Low-frequency coefficients of kth audio frame are denoted formally by $\hat{C}_k = \{\hat{c}_{ki} \mid 0 \leq i < L/2\}$, and their mean value is calculated by $\bar{\hat{c}}_k = \frac{1}{L/2}\sum_{i=1}^{L/2} \hat{c}_{ki}$. The watermark extraction rule is as follows:

$$w'_k = \begin{cases} 1 & \text{if } \bar{\hat{c}}_k - \lfloor \bar{\hat{c}}_k/\Delta_k \rfloor \cdot \Delta_k \geq \Delta_k/2 \\ 0 & \text{if } \bar{\hat{c}}_k - \lfloor \bar{\hat{c}}_k/\Delta_k \rfloor \cdot \Delta_k < \Delta_k/2 \end{cases} \tag{8}$$

In the watermarking algorithm, each audio frame has only a parameter Δ_k. In addition, when watermark embedding, if watermarking capacity provided by an audio signal is larger than watermark bits, watermark signal will be repeatedly embedded into the audio signal. When watermark extraction, a simple voting strategy is used to enhance robustness in the repeatedly embedding case.

3.3. **Audio watermarking framework.** In this work, the presented audio watermarking framework will apply the MOPSO-fs to optimize watermarking parameters of the audio watermarking algorithm so that its performance objectives can be optimally balanced. The audio watermarking framework can provide multiple optimal solutions, which allow us to select most suitable solution to construct an optimal watermarking scheme according to practical application demand. For clarity, we suppose the multi-objective optimal watermarking problem to be solved has K performance objectives, where $(K-1)$ attack methods are considered here. Thus, first objective is the fidelity while other objectives are the robustness against $(K-1)$ attacks respectively. The main components of the audio watermarking framework are summarized as follows.

3.3.1. *Representation.* In order to express the particles in swarm, we should determine its dimension firstly. However, the dimension D of the particle usually depends on the number of the parameters in watermarking algorithm. Therefore, dimension of particle is $D = N$ since each audio frame has only a parameter Δ_k. Formally, position and velocity vectors of mth particle (denoted by P_m) are $X_m = (x_m^1, x_m^2, \ldots, x_m^D)$ and $V_m = (v_m^1, v_m^2, \ldots, v_m^D)$, respectively.

3.3.2. *Initialize swarm.* Suppose the size of swarm is M and the size of repository is R. Initially, M particles are randomly generated in swarm. Note that velocity and position vectors of each particle should be initialized such that each candidate solution (particle) can be in the feasible searching space, i.e., $\beta_i \leq x_i \leq \gamma_i$. The memory $pbest_m$ of each particle is filled with its current position. And then, the repository is filled with all the non-dominated particles. Fitness sharing f_{sh}^m is calculated for each of particles in the repository.

3.3.3. *Update the velocity and position of each particle.* According fitness sharing of particles in the repository, some particles from the repository are chosen as leaders. Note that the leaders are chosen according to roulette wheel selection, i.e., particles with higher levels of fitness are likely to be selected. This will allow them to explore places less explored in the search space. And then, a particle from these leaders is selected randomly, whose position is regarded as *gbest*. The velocity and position vectors of each particle are updated according to Equations (1) and (2), respectively. Here, the inertia weight w in Equation (1) employs a linearly decreasing method [19].

3.3.4. *Accomplish watermark embedding.* Each particle P_m in swarm, whose position vector is $X_m = (x_m^1, x_m^2, \ldots, x_m^D)$, corresponds to a group of watermarking parameters, $x_m^1, x_m^2, \ldots, x_m^D$. For each particle, we execute watermark embedding algorithm on original audio signal to produce the watermarked audio signal according to the corresponding watermark parameters. The swarm has M particles, thus M watermarked audio signals are generated in all. And then, for each watermarked audio signal, we calculate its PSNR value by using Equation (4). Thus, first objective value of the particle is $f_{m1} = f_{m1}(x_m^1, x_m^2, \ldots, x_m^D) = \mathrm{PSNR}_m(x_m^1, x_m^2, \ldots, x_m^D)$.

3.3.5. *Accomplish watermark extraction.* For each watermarked audio signal, we attack it by using $(K-1)$ attack methods, thus, $(K-1)$ attacked audio signals are generated. As a result, the M particles have $M \times (K-1)$ attacked audio signals in all. For each attacked audio signal, we execute watermark extraction algorithm on it to extract watermark signal, and then calculate NC value of the extracted watermark signal. Therefore, each particle corresponds to $(K-1)$ NC values of the extracted watermark signals, i.e., latter $(K-1)$ objective values of the particle are

$$f_{m2} = f_{m2}(x_m^1, x_m^2, \ldots, x_m^D) = \mathrm{NC}_{m1}(x_m^1, x_m^2, \ldots, x_m^D), \ldots\ldots,$$
$$f_{mK} = f_{mK}(x_m^1, x_m^2, \ldots, x_m^D) = \mathrm{NC}_{m(K-1)}(x_m^1, x_m^2, \ldots, x_m^D).$$

3.3.6. *Update repository.* The repository is updated according to dominate and fitness sharing criteria. In the case where the repository is no full of non-dominated particles, if the particles dominate those particles inside the repository, they will be inserted whereas all solutions dominated will be deleted. In the case where the repository is full of non-dominated particles, if a particle non-dominated by any in the repository is found, their fitness sharing is calculated. If it is better than the worst fitness sharing in the repository, then the new particle replaces the one with the worst fitness sharing. The fitness sharing of all particles is updated when a particle is inserted in the repository or deleted from the repository. Thus, the repository is maintained as the Pareto front found so far.

3.3.7. *Update memory.* For each particle in swarm, if its current position dominates the one stored in its memory $pbest_m$, the one in memory will be replaced by the current one.

3.3.8. *Termination criteria.* The multi-objective optimization process described above is repeated until a terminated criteria is met. Usually, the terminated criteria can be a user-defined maximum iteration number or finding no other new non-dominated solution in a predefined number of successive iteration. In our experiments, the former is considered as our terminated criteria.

3.3.9. *Make decision for designing optimal watermarking scheme.* The Pareto optimal set generated by MOPSO-fs consists of multiple solutions, each of which can optimally balance all conflicting objectives of the watermarking algorithm. The property of the Pareto optimal set provides a flexibility for designing optimal watermarking scheme. Consequently, we can select a most suitable solution from the Pareto optimal set to construct an optimal watermarking scheme according to practical demand of watermarking application.

4. **Experiments and Discussions.**

4.1. **Data set.** Unlike image watermarking, audio watermarking has no standard audio data set, which is open and commonly used. Therefore, in order to measure capability of the proposed audio watermarking framework, we choose four kinds of music with different styles as our audio data set, including Jazz, Classical music, Blue and Rock-and-roll. The reason of using digital music is that digital musical files are broadly distributed and used in Internet. For every musical style, we collect 10 pieces of music as our test audio signals. These audio signals are about 2 minutes in length, mono, 16 bits/sample, 44.1kHz sampling rates.

In order to evaluate performance effects of different audio lengths for the audio watermarking framework, we randomly cut out four audio segments from each of above 40 audio signals, respectively. As a result, 160 audio signals are generated in all, and their lengths are 20, 40, 60 and 80 seconds respectively. These generated audio signals are added into the audio data set. On the other hand, in order to measure effects of different sampling rates, we convert each of above 40 audio signals to audio signals with 8kHz, 11.02kHz and 22.05kHz sampling rates, respectively, thus we obtain 120 audio signals, which are also contained in the data set.

4.2. **Setup.** Generally, the proposed audio watermarking framework has two classes of watermarking parameters: one is the parameter of the used watermarking algorithm; the other is the parameter of MOPSO-fs algorithm. The watermarking parameters of the used watermarking algorithm will be automatically determined by the proposed audio watermarking framework, so we only give parameter setup of MOPSO-fs algorithm below. According to successful results in many practical applications of MOPSO-fs algorithm, some priori parameters in experiments are empirically chosen as follows.

- According to experimental results from Salazar-Lechuga [16], population size and repository size in MOPSO-fs algorithm are empirically set to be $M = 100$ and $R = 50$ respectively, and let $\sigma_{share} = 2.0$. Since larger weight w tends to encourage global exploration and conversely smaller initial inertia weight encourages local exploitations, Shi and Eberhart [19] has suggested to vary w linearly from 0.9 to 0.4 over the course of the run. We allow w to vary between 0.9 to 0.4 linearly since it give a fast convergence over 100 iterations.
- The dimension D of particle's position and velocity vectors in PSO depends on length N_1 of audio signal and length L of audio frame. Therefore, $D = \lfloor N_1/L \rfloor$. In experiments, we set the length of audio frame to be $L = 1024$. Moreover, Daubechies-1 wavelet basis is used and one-level discrete wavelet transform is performed.

Usually, selection of attack methods can be determined according to practical demands. In experiments, we select five commonly used attack methods to illustrate the capability of the proposed audio watermarking framework, including additive noise, lossy compression (MP3), low-pass filtering, re-sampling and re-quantizing. Of course, we can use other attack methods and the audio watermarking framework is also available for them.

The watermark signal used in experiments is a binary random sequence generated by a Gaussian distribution with zero mean and unit variance. The length of the watermark signal equals to the number of audio frames in the audio signal used, i.e., $N_2 = \lfloor N_1/L \rfloor$. We denote the proposed audio watermarking framework by MOPSO below.

4.3. **Optimization results.** In experiments, we execute MOPSO over our data set and finally generate its Pareto fronts, i.e., Pareto optimal set. Note that each solution in Pareto optimal set is actually a group of optimal watermarking parameters. Therefore, for any audio signal in data set, we can generate its watermarked audio signal according

Table 1. The optimization results of MOPSO

Audio Signal	PSNR Range (dB)	Number of Solutions in Pareto Set	Additive Noise (NC)	MP3 (64kb) (NC)	Low-pass Filtering (NC)	Re-sampling (NC)	Re-quantization (NC)
Jazz	49∼55	49	0.97∼0.99	0.98∼0.99	0.94∼0.96	0.93∼0.95	0.99∼1
Classical music	49∼55	48	0.97∼0.99	0.98∼0.99	0.94∼0.96	0.93∼0.95	0.99∼1
Blues	49∼55	48	0.97∼0.99	0.98∼0.99	0.94∼0.96	0.93∼0.95	0.99∼1
Rock-and-roll	49∼55	49	0.97∼0.99	0.98∼0.99	0.94∼0.96	0.93∼0.95	0.99∼1

to each solution in the Pareto optimal set. Table 1 shows optimization results of MOPSO, where NC value under each attack is mean value over all audio signals with same musical style. From Table 1, we can see that PSNR values of all watermarked audio signals are in range 45dB $\sim$ 55dB, which means good fidelity. Furthermore, these results also indicate that MOPSO has strong robustness against noise, lossy compression (MP3), low-pass filtering, re-sampling and re-quantization.

From above experimental results, we can observe such a fact that MOPSO has approximately equal NC values over different style audio. This means that performance effect of different style audio for MOPSO is little. Nevertheless, from their Pareto optimal solutions, we observe that their watermarking parameters are obviously different. This fact indicates that for different style audio, the proposed audio watermarking framework can adaptively find their optimal parameters such that they have similar performances.

Unlike existing single-objective watermarking schemes that generate only a solution, the multi-objective watermarking scheme can generate a Pareto solution set that consists of multiple solutions. From Table 1, we can see that Pareto set contains about 47 $\sim$ 49 solutions which have different performances, and each of them corresponds to a group of optimal watermarking parameters. Since different watermarking applications may have different performance demands, the Pareto set provides the flexibility to select a suitable group of watermarking parameters from it.

4.4. **Compare with existing single-objective watermarking schemes.** The proposed audio watermarking framework uses MOPSO-fs to determine optimal watermarking parameters automatically and obtains a Pareto optimal set in which each group of parameters can guarantee optimal balance between fidelity and robustness. However, the existing watermarking schemes based on GA or PSO are essentially single-objective watermarking schemes because they convert the optimal watermarking problem to single-objective optimization problem by employing a weighted sum fitness function [7, 8, 9, 10, 11]. If we employ the PSNR and NC values to measure fidelity and robustness respectively, the weighted sum fitness function used in the single-objective watermarking schemes can be expressed by

$$fitness = \mathrm{PSNR} + \sum_{i=1}^{K} \lambda_i \mathrm{NC}_i \tag{9}$$

where λ_i is weighted factor. In their works, in order to guarantee the balance between fidelity and robustness, such a heuristic knowledge is used to determine the weighted factor λ_i, i.e., first item in Equation (9) should approximately equal to its second item. For simplicity, suppose these λ_i be equal and only five attack methods are considered. According to the heuristic knowledge, the weighted factors in the single-objective watermarking schemes can be determined as follows. As we know, for many audio watermarking schemes, PSNR is in range 45 $\sim$ 55 and NC is in range 0 $\sim$ 1. Therefore, we can set

Table 2. The comparison results between multi-objective framework and single-objective scheme over Jazz style audio (mean NC value)

Methods	PSNR (dB)	Additive Noise (NC)	MP3 (64kb) (NC)	Low-pass Filtering (NC)	Re-sampling (NC)	Re-quantization (NC)
MOPSO	53.26	0.9799	0.9837	0.9468	0.9415	1
SOGA	52.55	0.9543	0.9417	0.9295	0.9271	1

$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 10$. In addition, unlike MOPSO-fs that generates multiple solutions, single-objective optimization technique obtains only a solution.

In experiments, in order to compare performances of the proposed audio watermarking framework and the existing single-objective watermarking schemes, we implement a single-objective watermarking scheme based on GA by employing same watermarking algorithm and using the fitness function in Equation (9), which is denoted by SOGA. In GA, population size is 30, and chromosome uses binary string encoding and ranking selection, while the probabilities of crossover and mutation are chosen to be 0.7 and 0.005 respectively. Let largest generation number be 500. Table 2 gives mean performance results of MOPSO and SOGA over Jazz style audio, where the results of MOPSO correspond to a solution in its Pareto set. The comparison results clearly show that the proposed audio watermarking framework obviously exceeds the single-optimization watermarking scheme. In fact, we also obtain similar experimental results over other audio signals in our data set.

As stated above, not only fidelity and robustness are conflicting with each other, but also the robustness against different attacks are conflicting with each other. Form the view of multi-objective optimization, the existing single-objective watermarking schemes failed to optimally balance the conflicting objectives. However, multi-objective watermarking scheme can generate a Pareto solution set under dominance meaning such that the conflicting objectives can be balanced optimally. These comparison results indicate the following two facts. Firstly, compared with the existing single-objective watermarking scheme, the proposed audio watermarking framework can make the used watermarking algorithm to hold its best performance. Secondly, the weighted factors determined according to above heuristic knowledge may not obtain its optimal watermarking parameters for a watermarking algorithm, so the single-objective watermarking schemes failed to attain or approach it performance upper limits. In addition, it is difficult task to determine optimal weighted factors of the weighted fitness function in single-objective optimization.

4.5. **Compare with other methods.** In the following, we will compare results of the proposed audio watermarking framework with the performances of three existing watermarking schemes. The three watermarking schemes are described as follows.

- First watermarking scheme is from Chen and Lin [18]. When we implement the scheme, its original parameters are employed. The scheme is denoted by WMA.
- Second watermarking scheme is the scheme using artificial neural network proposed in Wang et al. [20]. The parameters in original literature are employed to implement the scheme. The scheme is denoted by WMA-ANN.
- Third watermarking scheme is the scheme using support vector regression proposed in Xu et al. [21]. The parameters in original literature are employed to implement the scheme. The scheme is denoted by WMA-SVR.

In experiments, performance comparisons are completed by two groups of experiments. First group compares MOPSO with WMA. The purpose of the group comparison is

TABLE 3. The comparison results of several watermarking schemes over Jazz style audio

Methods	PSNR (dB)	Additive Noise(25) (NC)	MP3 (64kb) (NC)	Low-pass Filtering (NC)	Re-sampling (NC)	Re-quantization (NC)	Noise + MP3 (NC)	Noise + Low-pass (NC)
MOPSO *	52.57	0.9871	0.9891	0.9532	0.9437	1	0.9586	0.9345
MOPSO **	53.26	0.9799	0.9837	0.9468	0.9415	1	0.9557	0.9321
WMA	52.16	0.9428	0.9153	0.9136	0.9081	0.9678	0.9026	0.9015
WMA-ANN	50.63	0.9532	0.9049	0.9239	0.9052	0.9094	0.9015	0.8991
WMA-SVR	50.85	0.9639	0.9332	0.9319	0.9175	1	0.9087	0.9063

TABLE 4. The comparison results of several watermarking schemes over classical style audio

Methods	PSNR (dB)	Additive Noise(25) (NC)	MP3 (64kb) (NC)	Low-pass Filtering (NC)	Re-sampling (NC)	Re-quantization (NC)	Noise + MP3 (NC)	Noise + Low-pass (NC)
MOPSO *	52.54	0.9869	0.9888	0.9531	0.9436	1	0.9583	0.9344
MOPSO **	53.23	0.9786	0.9835	0.9466	0.9413	1	0.9556	0.9319
WMA	52.14	0.9426	0.9151	0.9135	0.9082	0.9679	0.9025	0.9013
WMA-ANN	50.61	0.9434	0.8951	0.9177	0.9012	0.9051	0.9012	0.8987
WMA-SVR	50.81	0.9546	0.9255	0.9193	0.9106	1	0.9076	0.9059

TABLE 5. The comparison results of several watermarking schemes over Blues style audio

Methods	PSNR (dB)	Additive Noise(25) (NC)	MP3 (64kb) (NC)	Low-pass Filtering (NC)	Re-sampling (NC)	Re-quantization (NC)	Noise + MP3 (NC)	Noise + Low-pass (NC)
MOPSO *	52.53	0.9868	0.9890	0.9529	0.9435	0.9997	0.9581	0.9342
MOPSO **	53.24	0.9785	0.9837	0.9468	0.9412	1	0.9555	0.9317
WMA	52.12	0.9424	0.9149	0.9133	0.9081	0.9676	0.9023	0.9011
WMA-ANN	50.59	0.9507	0.9011	0.9215	0.9034	0.9077	0.9011	0.8985
WMA-SVR	50.76	0.9617	0.9314	0.9286	0.9165	1	0.9083	0.9062

whether the proposed audio watermarking framework effectively improves the performance of conventional watermarking algorithm. Another group compares MOPSO with WMA-ANN and WMA-SVR. The WMA-ANN and WMA-SVR are two recently proposed audio watermarking schemes based on artificial neural network and support vector machine respectively, which exhibit promising performance results. Tables 3-6 show comparison results of these watermarking schemes over Jazz, Classical music, Blue and Rock-and-roll style audio, respectively. In every table, we provide two groups of optimization results of MOPSO, which correspond to two solutions (two groups of optimal watermarking parameters) in its Pareto set.

For the comparison results of MOPSO with WMA, we can observe that performance results of MOPSO evidently exceed ones of WMA. These comparison results indicate the following two facts. On the one hand, the proposed audio watermarking framework can effectively improve performance of conventional watermarking algorithm, guaranteeing optimal balance between its conflicting objectives. On the other hand, watermarking parameters determined empirically may not make conventional watermarking scheme to be

Table 6. The comparison results of several watermarking schemes over Rock-and-Roll style audio

Methods	PSNR (dB)	Additive Noise(25) (NC)	MP3 (64kb) (NC)	Low-pass Filtering (NC)	Re-sampling (NC)	Re-quan-tization (NC)	Noise + MP3 (NC)	Noise + Low-pass (NC)
MOPSO *	52.58	0.9873	0.9892	0.9531	0.9438	1	0.9587	0.9344
MOPSO **	53.29	0.9791	0.9839	0.9467	0.9418	1	0.9556	0.9323
WMA	52.18	0.9429	0.9155	0.9135	0.9083	0.9677	0.9028	0.9016
WMA-ANN	50.68	0.9475	0.9083	0.9274	0.9095	0.9124	0.9019	0.8998
WMA-SVR	50.89	0.9586	0.9377	0.9339	0.9216	1	0.9072	0.9078

optimal for different audio signals, that is, it is a difficult task to determine optimal watermarking parameters of conventional watermarking scheme according to a person's experience. However, the proposed audio watermarking framework can effectively avoid the problem because its optimal watermarking parameters can be automatically determined by the audio watermarking framework. From Tables 3-6, we can see that performance of MOPSO evidently exceeds ones of WMA-ANN and WMA-SVR. The comparison result indicates that the proposed audio watermarking framework can make conventional watermarking algorithms to attain or approach their performance upper limits.

5. **Conclusions.** Aiming at the conflicting multi-objectives of optimal audio watermarking problem, we present a novel audio watermarking framework based on multi-objective particle swarm optimization, which automatically determine its optimal watermarking parameters. In contrast, conventional watermarking algorithms determine their watermarking parameters empirically or by experiment, so they do not balance their conflicting objectives optimally. In addition, we also observe that the watermarking framework has approximately equal performances over audio with different musical styles whereas their optimal watermarking parameters are quite different. This result indicates that though audio with different musical style have different local features, the proposed audio watermarking framework can adaptively determine optimal watermarking parameters such that it can hold similar performances over these audio.

Compared with the existing single-objective watermarking schemes, the proposed audio watermarking framework can obviously improve the performances of conventional watermarking algorithms. The comparison results also indicate that the heuristic knowledge used in the existing single-objective watermarking schemes may not get optimal parameters for a watermarking algorithm. Furthermore, the proposed audio watermarking framework can avoid the difficulty of determining optimal weighted factor in the existing single-objective watermarking schemes.

Finally, note that though we consider only five common attacks in experiments, the proposed audio watermarking framework is also suitable for other attack methods. In addition, limiting to the audio watermarking algorithm used in this paper, which does not resist desynchronization attacks, we do not consider the desynchronization attack methods in experiments. As we know, it is a challenging issue to resist desynchronization attacks, our further work is to design a new audio watermarking algorithm to solve the challenging issue under the proposed audio watermarking framework.

Acknowledgements. This work was partially supported by the National Natural Science Foundation of China (No. 61170030), Research Fund of Sichuan Key Technology Research and Development Program (No. 2012GZ0019), Open Research Fund of Sichuan

Key Laboratory of Intelligent Network Information Processing (No. SZJJ2012-030), Importance Project Foundation of the Education Department of Sichuan province (No. 12ZA163), and Importance Project Foundation of Xihua University (No. Z1122632), China.

REFERENCES

[1] I. J. Cox and M. L. Miller, The first 50 year of electronic watermarking, *Journal of Applied Signal Processing*, vol.2002, no.2, pp.126-132, 2002.

[2] H. Peng, X. Wang, W. Wang, J. Wang and D. Y. Hu, Audio watermarking approach based on audio features in multiwavelet domain, *Journal of Computer Research and Development*, vol.47, no.2, pp.216-222, 2010.

[3] H. Peng, J. Wang and Z. Zhang, Audio watermarking scheme robust against desynchronization attacks based on kernel clustering, *Multimedia Tools and Applications*, vol.62, no.3, pp.681-699, 2013.

[4] J. Wang, H. Peng and P. Shi, An optimal image watermarking approach based on a multi-objective genetic algorithm, *Information Sciences*, vol.181, no.24, pp.5501-5514, 2011.

[5] H. Peng, J. Wang and W. Wang, Image watermarking method in multiwavelet domain based on support vector manchines, *Journal of System and Software*, vol.83, no.8, pp.1470-1477, 2010.

[6] C.-C. Lai, H.-C. Huang and C.-C. Tsai, A digital watermarking scheme based on singular value decomposition and micro-genetic algorithm, *International Journal of Innovative Computing, Information and Control*, vol.5, no.7, pp.1867-1874, 2009.

[7] C. H. Huang and J. L. Wu, A watermark optimization technique based on genetic algorithms, *The SPIE Electronic Imaging*, San Jose, CA, USA, pp.516-523, 2000.

[8] C. S. Shieh, H. C. Huang, F. H. Wang and J. S. Pan, Genetic watermarking based on transform domain techniques, *Pattern Recognition*, vol.37, no.3, pp.555-565, 2004.

[9] P. Kumsawa, K. Attakitmongcol and A. Srikaew, A new approach for optimization in image watermarking by using genetic algorithm, *IEEE Transactions on Signal Processing*, vol.53, no.12, pp.4707-4719, 2005.

[10] C. H. Huang and J. L. Wu, Fidelity-guaranteed robustness enhancement of blind-detection watermarking schemes, *Information Sciences*, vol.179, no.6, pp.791-808, 2009.

[11] F. M. Meng, H. Peng, Z. Pei and J. Wang, An adaptive image watermarking scheme based on support vector machine and genetic algorithm, *Pattern Recognition and Artificial Intelligence*, vol.22, no.2, pp.312-317, 2009.

[12] H. Peng and J. Wang, Optimal audio watermarking scheme using genetic optimization, *Annals of Telecommunications*, vol.66, no.5-6, pp.307-318, 2011.

[13] Y.-R. Wang, W.-H. Lin and L. Yang, An intelligent watermarking method based on particle swarm optimization, *Expert Systems with Applications*, vol.38, no.7, pp.8024-8029, 2011.

[14] R. C. Eberhart, J. Kennedy and Y. H. Shi, *Swarm Intelligence*, San Mateo, Morgan Kaufmann, CA, USA, 2001.

[15] J. Kennedy and R. C. Eberhart, Particle swarm optimization, *Proc. of IEEE Int. Conf. Neural Networks*, Perth, Australia, vol.4, pp.1942-1948, 1995.

[16] M. Salazar-Lechuga and J. E. Rowe, Particle swarm optimization and fitness sharing to solve multi-objective optimization problems, *Proc. of Congr. Evol. Comput.*, pp.1204-1211, 2005.

[17] S. K. Goudos, Z. D. Zaharis, D. G. Kampitaki, I. T. Rekanos and C. S. Hilas, Pareto optimal design of dual-band base station antenna arrays using multi-objective particle swarm optimization with fitness sharing, *IEEE Transactions on Magnetics*, vol.45, no.3, pp.1522-1525, 2009.

[18] L. H. Chen and J. J. Lin, Mean quantization based image watermarking, *Image and Vision Computing*, vol.21, no.8, pp.717-727, 2003.

[19] Y. Shi and R. C. Eberhart, A modified particle swarm optimizer, *Proc. of IEEE World Congr. Evol. Comput. Intell.*, pp.69-73, 1998.

[20] C. Wang, X. H. Ma, X. P. Cong and F. L. Yin, An audio watermarking scheme with neural network, *International Symposium on Neural Networks, LNCS*, vol.3497, pp.795-800, 2005.

[21] X. J. Xu, H. Peng and C. Y. He, DWT-based sudio watermarking using support vector regression and subsampling, *International Workshop on Fuzzy Logic and Applications, LNAI*, vol.4578, pp.136-144, 2007.

[22] H. Peng, B. Li, X. Luo, J. Wang and Z. Zhang, A learning-based audio watermarking scheme using kernel Fisher discriminant analysis, *Digital Signal Processing*, vol.23, no.1, pp.382-389, 2013.