

Watermark Synchronization: Perspectives and a New Paradigm

Gaurav Sharma and David J. Coumou

Dept. of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627-0126

Email: {gsharma,coumou}@ece.rochester.edu

Abstract—Synchronization is one of the most challenging elements of a watermarking system. In this paper, we survey and classify methods employed for watermark synchronization and highlight some inherent difficulties in synchronization that arise due to the nature of multimedia signals. We then propose a framework that addresses these hurdles in practical applications. Our framework utilizes feature-based embedding in concert with error correction codes capable of handling not only substitutions caused by various perturbations but also insertion and deletion events caused by erroneous feature estimates. We present a practical scheme for speech watermarking based on the framework. Experimental results show that the proposed methodology indeed enables synchronization in scenarios where a mismatch in estimated features between embedder and receiver would otherwise cause synchronization loss. We explore connections of the framework with recent theoretical analyses of watermark synchronization.

I. INTRODUCTION

After more than a decade of multimedia watermarking development, watermark synchronization remains a vexing issue. Synchronization is an essential element of every digital communication system and has been extensively researched in that context. In watermarking/data-hiding applications, however, synchronization poses unusual and particularly challenging new problems because the primary goal in these systems is not the communication of the watermark data but the communication of the multi-media information with minimal or no perceptual degradation. The communication of the embedded data is a secondary objective that, nonetheless, is often required to be robust against signal processing operations that do not significantly degrade perceptual quality. A variety of watermarking schemes have been proposed to facilitate synchronization at the watermark receiver. Typically, methods are designed to be robust against a specific set of operations such as rotation, scaling, and translation, or some combination thereof, and have varying levels of success.

In this paper, our goals are fourfold: First, we classify several of the existing watermark synchronization methods based on their underlying principles and highlight their advantages and shortcomings. Next we advocate specifically on behalf of feature-based embedding techniques and present a new synchronization paradigm for this class of methods that overcomes their main weakness by combining them with error correction codes capable of recovering from insertions, deletions, and substitutions. Then we show that the proposed methodology can indeed enable synchronization in practice by developing a speech watermark that forms a preliminary

implementation in our new paradigm. Finally, we relate our practical approach to recent theoretical analyses of watermark synchronization.

II. WATERMARK SYNCHRONIZATION: A TAXONOMY

Depending on the watermarking method, synchronization errors may arise from a variety of channel degradations corresponding to both benign and malicious processing operations. Common watermark embedding methods are assumed to be conformal with the inherent geometry of the image. This is the case, for example, for spatial/temporal domain spread-spectrum watermarking for images and audio, and also transform domain methods (to a lesser extent). For these methods the most common and challenging source of de-synchronization errors is the class of "geometric attacks" that correspond to a change in the underlying geometry of the signals through modifications of the temporal/spatial axes. This includes common operations of scaling, cropping, translation, shearing, and rotation (the latter 2 apply to 2-D signals) as well as local warping of the signal geometry. Stirmark [1] provides a very effective tool for simulation of these common attacks. An alternative class of watermark embedding methods avoid several of these synchronization problems by requiring no conformance to the signal geometry for the watermark. Examples in this class are methods that embed watermark information in signal features, histograms, or moments. These methods can, however, also suffer from de-synchronization if the procedures used at the embedder and receiver are not entirely repeatable. In the following, we classify proposed remedies for synchronization considering methods in both classes mentioned above.

For non-oblivious watermarking systems in which the original may be assumed to be available at the receiver, synchronization can be achieved by registering the received signal to the original and using the registered version of the signal for watermark detection [2], [3]. Non-oblivious methods of detection are, however, unsuitable in most applications involving large multi-media databases, and we therefore focus entirely on oblivious watermarking systems in which the original cover signal is not available at the receiver.

For oblivious watermarking, commonly employed methods for watermark synchronization may be classified into four major categories: (1) methods that embed information in geometrically invariant transforms (GIT), (2) recovery assisted methods (RAM), (3) Feature-Based (FB) embedding methods, and (4) methods using exhaustive search (ES) at the detector.

- 1) **Geometric Invariant Transform (GIT):** The GIT class of methods embed the watermark information in a signal transform domain that is invariant to specific geometric transformations and thereby ensure that synchronization is not lost when the signal is subjected to these transformations. A specific example is embedding in the Fourier-Mellin transform for 2-D images that provides invariance to global rotation, scaling and translation [4]. A variety of other closely related transforms have also been utilized for this purpose [5], [6]. A limitation of these methods is that they only handle the specific class of transformations for which they are designed and typically do not directly handle other operations such as cropping and locally-varying geometric transformations such as the random bending attack in StirMark [1].
- 2) **Recovery Assisted Methods (RAM):** The RAM class of methods incorporate information in the watermarked image that allows estimation of global geometric transformations applied to the signal after watermark embedding. This allows the receiver to invert the geometric transformation and thereby achieve synchronization, provided the estimated parameters are accurate. Methods in this category may further be grouped as structured watermark (SWM) and template based (TB) techniques.

The *structured watermark (SWM)* class typically utilizes a repetitive pseudo-random noise pattern for watermark embedding with suitably chosen periodicity. If the watermarked image undergoes a geometric transformation between the embedder and receiver this periodicity is altered. In SWM the altered periodicity is determined at the receiver by performing an auto-correlation computation that reveals the repetitions as spikes in the auto-correlation. Using these estimates with the known embedding periodicity, the receiver can estimate the transformation parameters and undo these transformations. These techniques have been utilized with some success in image watermarking [9], [10].

Template based (TB) methods for watermark synchronization are closely related to the SWM methods in their operating principles. A minor difference is that typically the watermark payload in these methods is independent of the synchronization "template". The synchronization template may be embedded in the spatial or frequency domain [12], [13], [14] and once again enables estimation of global transformations applied to the signal between the embedder and the receiver.

Both SWM and TB methods are subject to removal attacks [15] that attempt to remove the correlation peaks/template. They also provide rather limited resilience against locally varying geometric transformations [11] and can be misled by an active attacker that adds another SWM/template to confuse the receiver. The SWM technique has an advantage over the TB method in that the power need not be split between the watermark and the template.

- 3) **Feature-Based (FB) methods:** The FB class may

be viewed as *second generation* watermarking algorithms [16] that evolve from watermark embedding in pixels toward embedding using semantically meaningful features of the multi-media signals either for the purpose of synchronization or for embedding itself. Common examples of feature extraction methods used in this application include corner detection [17], clustering and multi-resolution analysis [16]. Feature-based methods have received rather limited attention in main-stream watermarking literature. One reason for this is that despite their significant promise, the FB methods have faced very significant implementation hurdles since accurate and repeatable estimation of multi-media signal features is in itself a challenging research problem [18]. A mis-match in estimated features between the embedder and receiver, typically causes catastrophic de-synchronization. Therefore even in the absence of attacks the methods have hitherto failed ungracefully.

- 4) **Exhaustive Search (ES):** For spread-spectrum watermark embedding where the detector is required to make a binary decision corresponding to the presence or absence of a watermark, an alternative method for achieving watermark synchronization is to perform ES. The method performs a watermark detection over all possible geometric distortions and declares the watermark present if a positive detection is obtained postulated on some geometric distortion. The method is obviously practically feasible only when the class of geometric distortions can be parameterized by a reasonably small set of parameters. This limits the applicability of ES. Typically, the ES method also suffers from an undesirable effect that it can increase the probability of false alarm[19].

The classification presented above is not exhaustive and numerous other synchronization methods have been proposed. Combinations of methods listed may also be utilized in practice. Despite this large body of work, no widely acceptable and secure solution exists for watermark synchronization.

III. FEATURE-BASED EMBEDDING: PROS AND CONS

Among the synchronization methods presented above, the feature-based (FB) methods are among the least mature. Given the complexity of these techniques, analytic end-to-end system evaluation for estimating bit-error rates or capacity is usually infeasible and only limited analysis may be applied to their individual modules [18]. A discrepancy in estimated features between the embedder and the receiver causes a catastrophic de-synchronization for FB methods, often even under relatively benign channel conditions of non-malicious processing. Thus unlike other embedding methods where synchronization is viewed as an additional and separate requirement, these methods make it a pre-requisite for basic operation. Empirical evaluation has therefore also been hampered for want of robust synchronization methods.

The lack of an analytical framework combined with the absence of a practically viable implementation has ensured that FB watermarking methods have received rather limited

attention thus far. This, however, does not detract from their numerous advantages, which we elaborate next:

- Under global geometric transformations, well-chosen features will all undergo the “same” change in geometry.
- If the features are also semantically meaningful, they have resilience to lossy compression operations that attempt to remove irrelevant information while preserving perceptually/semantically important information.
- Localization based on features can also be more meaningful since it can be related to semantics and may therefore better withstand processing at a higher semantic level, such as copying of an object from one image into another.
- The use of features can also potentially allow better informed-embedding since the features can help identify regions unsuitable for embedding such as extremely smooth image regions with no discernible features or textured regions that may be well suited for embedding.
- Finally, the perceptual significance of the features can also limit the extent to which these are impacted by malicious attacks that modify the data with only small perceptual changes.

Thus, well-chosen features offer excellent resilience to a number of common degradations encountered in watermark channels and form a natural choice for achieving invariance and synchronization. The exact benefits realized, of course, depend on the embedding method adopted and the selected features. Some of the key desirable characteristics of features for FB watermark embedding have previously been listed in [16] as: a) invariance to noise/compression, b) covariance to geometric transformations, and c) localization. The first two are self-explanatory; localization requires that cropping or replacing the data in one region should not alter remaining features, i.e., the features should be locally computed. We assume that features satisfy these requirements and, based on the localization requirement, refer to them as “feature-points”. Arguing conversely, one can also see that non FB methods suffer from major shortcomings because the embedding method does not relate to signal semantics. These methods therefore offer very little or no resilience against higher level processing operations that operate at a semantic level. Thus *synchronization capabilities of non FB methods may be expected to be inherently limited* and we therefore strongly advocate additional research on FB watermarking methods. In current practice, however, FB methods are totally stymied by de-synchronization that occurs when feature estimation between the receiver and embedder is not repeatable. We next present a new synchronization and error-control coding paradigm that addresses this limitation.

IV. A NEW PARADIGM: FB EMBEDDING + IDS CODES

The fundamental problem in FB watermarking methods occurs due to a discrepancy in feature-points determined at the embedder using the original cover signal and those estimated using the signal available at the receiver, which may have undergone compression, noise addition, and possibly additional attacks. Typically, for localized features this discrepancy

is manifested in the form of some feature-points that are detected at the embedder but not at the receiver and vice-versa. Assuming that the data is embedded in the feature-points themselves,¹ this mismatch causes random insertions and deletions in the extracted data-stream. Combined with substitutions caused by other perturbations or noise, the system for embedding in features may therefore be viewed as a channel that provides data-transport albeit with insertions, deletions, and substitutions (IDS).

The insertions and deletions are the source of the debilitating de-synchronization errors that cannot be handled by error correction codes that are designed to correct substitutions (and form the most widely-studied class of codes). However, if one utilizes error-correction codes capable of handling IDS errors, these errors may also be corrected. Hence we propose a new paradigm for watermark synchronization that uses feature-based embedding and overlays the IDS channel provided by the feature-based embedding and extraction with IDS codes in order to recover from the synchronization errors. This synchronization framework is shown in Fig. 1. The basic data embedding and extraction technique is indicated as the block outlined with the solid border. As argued above, the method provides a data-transport channel from t to \hat{t} that potentially encounters IDS errors. To overcome these, watermark data m to be communicated is first encoded using a suitable error-correction code to obtain t and the received data \hat{t} is decoded using a corresponding decoder to obtain an estimated message \hat{m} . This provides the requisite synchronization and error recovery capability and is shown as the dotted block in Fig. 1. The resulting combination yields a novel watermarking framework that is able to achieve overall watermark synchronization despite loss of synchronization in the underlying data embedding method.

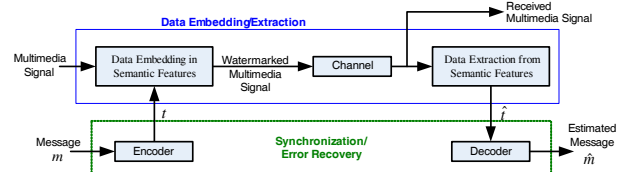


Fig. 1. Data embedding in features with synchronization using IDS codes.

Obviously, the viability of the proposed framework is dependent on the capability to develop suitable error correction codes. Given the nature of the IDS channel, practically useful classes of linear codes are not known for this application and the problem of code design in general is therefore a challenging one. To illustrate the viability of the framework with next present a practical watermark.

V. A SPEECH WATERMARK IN THE PROPOSED PARADIGM

In this section, we present a speech watermark that embodies the paradigm proposed in Section IV. We provide a pithy description of the application and results; the interested reader is directed to [20], [21] for a detailed description of the scheme, implementation details and additional results.

¹This assumption can be relaxed within this paradigm.

Our speech watermark with pitch-modification based embedding and IDS code based synchronization is schematically illustrated in Fig.2. The embedding and extraction of data into pitch is located on the right side of the figure and closely follows [22]. For the purpose of watermarking, we use pitch as the semantically meaningful feature. The selection of pitch as the feature for watermark embedding is motivated by the structure of most speech encoders that ensures the fidelity of the pitch information is maintained during compression.

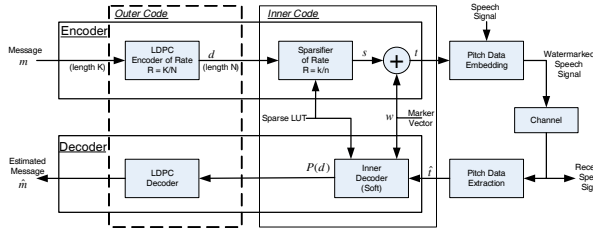


Fig. 2. Speech data-embedding by pitch modification and synchronization using an IDS code.

Data embedding employs scalar quantization index modulation (QIM) [23] of block pitch averages over non-overlapping intervals. A sequence of successive pitch intervals $\{p_i\}_{i=1}^J$ is modified as to embed an information bit in the average pitch by modifying it from p_{avg} to p'_{avg} , where $p_{avg} = \frac{1}{J} \sum_{i=1}^J p_i$ and $p'_{avg} = Q_b(p_{avg})$, respectively. $Q_b(\cdot)$ denotes the corresponding quantizer where b is the embedded bit from the information sequence t . The pitch modifications are incorporated into the speech signal using the pitch synchronous overlap add (PSOLA) algorithm [24].

The watermark receiver extracts embedded data from the speech signal by partitioning voiced segments into non-overlapping blocks of pitch intervals and computing average pitch values over J contiguous pitch intervals. For each block average, an estimated value of the embedded bit is computed as the index 0/1 from the quantizer $Q_b(\cdot)_{b=0}^1$ for a reconstruction value closest to the average pitch. This extraction process yields the estimate \hat{t} of the embedded data t .

As indicated in Section III, a mismatch in estimated features between the sending and receiving ends causes desynchronization. In data embedding by pitch modification, this occurs when estimates of voiced segments at the receiver differ from those at the embedder [22]. Multiple voiced segments at the embedder may coalesce into a single voiced segment at the receiver, or vice versa. In addition, relatively small voiced segments may be detected at one end and not the other. These types of mismatches result in insertion, deletion, and substitution (IDS) errors in \hat{t} yielding a loss of synchronization.

To resolve the synchronization problem between the watermark embedder and receiver, we employ the IDS error correcting codes recently proposed by Davey and MacKay [25]. The error correction scheme, also shown in Fig. 2, consists of an outer q -ary LDPC code concatenated with a nonlinear inner code that aids synchronization. The watermark encoder applies an inner and outer encoding processes to the q -ary (q

$= 2^k$ bits) message m .

The inner (nonlinear) code operates using a key-based pseudo random marker vector w that is known to both the embedder and receiver. The marker vector w , with minor perturbations, is embedded into the speech. The perturbations of w incorporate the watermark data to be communicated. At the receiver, the knowledge of w enables recovery of synchronization with some uncertainty. This is accomplished using an hidden-Markov-model (HMM)-based soft inner decoder that outputs symbol-by-symbol MAP probabilities for the q -ary symbols. The uncertainty arises due to channel induced IDS errors as well as the perturbations caused by data embedding. The outer q -ary LDPC code resolves this uncertainty and using the known values of w allows recovery of the perturbations introduced in w and thereby the embedded watermark data. Note that q -ary, instead of binary, codes offer a significant benefit not just from their improved error performance but also from the fact that they allow better synchronization [25].

VI. EXPERIMENTAL RESULTS

To evaluate the pitch-modification based speech watermark with synchronization, the method of the previous section was tested with sample speech files from a database available at [26]. For the pitch-based embedding, it was determined through (limited) empirical testing that using a QIM quantization size of $\Delta \leq 15$ resulted in an embedding distortion that was imperceptible. However, the choice of this step size did not provide immunity against synchronization loss. Fig. 3 shows a small region of the transmitted and received data streams plotted against an absolute time reference (corresponding to the pitch analysis windows). The stars on the figure represent locations at which a bit was embedded and the squares indicate locations at which the receiver guesses that a bit was embedded, with the ordinate (y) axis indicating the value of the bit as 0/1. It is clear that the embedding and extraction process introduces insertions, deletions and substitutions as labeled in the plot. In the absence of the outer error-correction layer of the proposed system, the desynchronization resulting from the insertion/deletion events yields a bit error rate that asymptotically tends to $\frac{1}{2}$.

Next the proposed system was evaluated. For this purpose, random $q = 16$ -ary messages m of length $K = 25$ were generated and encoded using a rate $\frac{1}{4}$ q -ary LDPC code. These were then mapped through a sparsifier to vectors of length $n = 1000$ bits added (modulo-2) to a pseudo-random marker vector w and embedded into speech files using the pitch based QIM embedding with step size $\Delta = 10$. The resulting files were subjected compression "channels" corresponding to: a) No compression, b) GSM-06.10 at 13 kbps, and c) AMR at 5.1 kbps. In all test cases investigated, it was found that the proposed system correctly recovered the embedded data provided sufficient number of iterations were allowed for the outer LDPC decoder.

To further examine the performance of the proposed method, Monte-Carlo simulations were performed and the percent of times that successful synchronization and data

recovery² was accomplished was determined. This percentage was computed for different values of the QIM embedding step size Δ for the case of no compression. The corresponding results are shown in Fig. 4. Note that the performance degrades with a decrease in QIM step size as expected. For $\Delta = 15$ no cases of failed synchronization/recovery were observed and for $\Delta = 10$ over 95% of the times synchronization was regained. These indicate that the method can successfully be utilized at QIM step sizes that are imperceptible with relatively low probability of unrecoverable error.

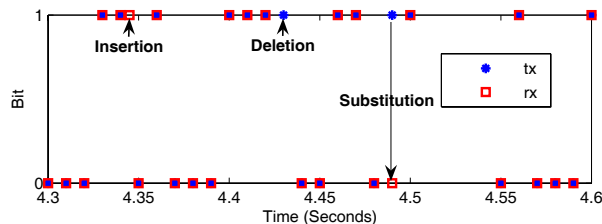


Fig. 3. IDS events in pitch-based embedding.

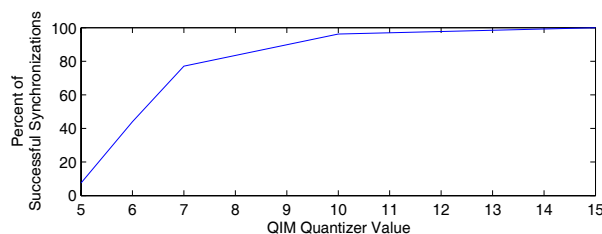


Fig. 4. Monte-Carlo simulation results illustrating dependence of synchronization performance on QIM step size.

VII. DISCUSSION

Some elements of the proposed watermark have strong similarity with the template-based (TB) and structured-watermark (SWM) methods. The marker vector w which enables the synchronization may be considered analogous to a template or along with the watermark information analogous to a SWM. However, the analogy is a rather loose one since the SWM and TB methods rely on an embedding conformal with the signal geometry in order to allow recovery of geometric distortion parameters whereas in the proposed scheme the feature-based embedding enables freedom from this requirement. Also, in scenarios where w is generated using a cryptographic key, it prevents detectors without access to the key from recovering the watermark data. Finally w does not include any tell-tale structure or repetition and is therefore immune to the template-removal attack. Thus the proposed framework does not suffer from any of the limitations of the SWM and TB class of methods. As we point out subsequently, in fact our proposed watermark is in fact a closer relative of the exhaustive search (ES) class of methods.

²In LDPC decoding operating at a reasonable rate and block length, a successful decoding typically implies correct data recovery (in other words undetectable error events are extremely rare). Therefore, percent successful synchronization is used instead of error rate after decoding.

The speech watermark that we present does have some limitations. Firstly, the embedding in pitch averages over windows provides a rather limited data rate. Combined with the rather low rate of the concatenated error-correction code for synchronization and recovery, this further limits the effective end-to-end rate of the overall system. Based on other recent work [27], we speculate that this may in fact be a necessary trade-off for synchronization and is probably a limitation of the general framework too. It is also worth noting that in some systems this low rate may not be a hindrance. For instance, depending on the application (e.g. fingerprinting) it may be sufficient if only a unique object identifier is embedded using a method resilient to de-synchronization, which may then be used to recover the original signal from a database and perform synchronization for non-oblivious detection of a more useful payload.

A second limitation is applicable specifically to the speech watermark embedding method presented above. While the method provides excellent resilience to low-bit-rate lossy compression (see Section VI), the pitch-based QIM embedding does not provide any resilience against deliberate scaling of the time axis³. This limitation, however, arises from the specific embedding method and does not apply to the framework in general. Alternate, well-designed FB embedding methods need not suffer from this limitation.

In a recent paper, Barni [28] analyzed exhaustive search (ES) and template-based (TB) watermarking in a simple but elegant setting. The results indicate that *if the size of the search space does not increase exponentially*, both ES and TB methods provide asymptotically good results. The ES method also provides superior detection performance over TB methods (as determined by ROC curves). The analysis for the ES methods has also been extended by Merhav in [29] where the embedder and detector have been optimized under a resource constrained detection scenario that uses only the energy and correlation. Once again, the asymptotic optimality of exhaustive search is established. Both analyses predicate their results on a polynomial size for the search space. For embedding methods conformal with the image geometry and general geometric distortions, this assumption will probably fail for most practical methods. However, one can immediately realize that *the size of the search space is immaterial and the complexity of the search (for the optimum) over the space is the real quantity of interest*. In particular, under suitable constraints dynamic programming may allow efficient search over an exponential space. This in fact the case for our speech watermark when it is suitably "de-featured" to a zero bit watermark, which corresponds to purely the detection of the presence or absence of w when no watermark data is embedded. The feature-based embedding reduces the size of the embedding domain; however, the search space remains exponential in the number of features. The HMM representation

³This corresponds to a valumetric scaling attack on the QIM embedder.

for the channel, however, enables us to do an efficient search⁴ over this exponential space using dynamic programming. In this context, it is worth noting that the marker vector w plays the same role as the spread spectrum watermark [28]. Our detector, however, differs from the ones used in [28], [29] in that it does not operate under the same resource constraints. Also note that in this reduced mode of operation for our scheme, the probability of false alarm encounters little or no increase because the LDPC error correction decoding validates the embedded data (and there are few undetected error events). The same is true for ES used in detection of cropping in authentication watermarks (e.g., see [30]), where the fact that the embedded information is a "signature" of the image data makes false positives a moot concern.

VIII. CONCLUSIONS

A synchronization paradigm was proposed for feature-based watermarking that brings us closer to exploiting the inherent benefits of these methods. We show that the fundamental synchronization problem in feature-based watermarking caused by discrepancies in features estimated at the embedder and receiver can be overcome through the use of error-correction codes capable of correcting insertions, deletions, and substitutions (IDS). The paradigm thus eliminates the requirement for perfect feature recovery at the receiver and paves the way for practical feature-based watermarking methods.

We demonstrated a speech watermark implemented in the proposed framework that uses pitch modification for data-embedding and the IDS-correcting codes proposed by Davey and MacKay [25]. Experimental results illustrate successful synchronization recovery in our system and show that the paradigm is powerful and practical and allows the speech watermark to be resilient to lossy compression using low bit rate speech CODECs.

With the example implementation of the synchronization paradigm, we have illustrated a viable method for addressing the vexing issue of synchronization in feature-based watermarking and anticipate that the paradigm will provide a much needed boost to similar synchronization methods for other feature-based methods.

REFERENCES

- [1] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Attacks on copyright marking systems," in *Information Hiding, Second International Workshop, IH'98*, Portland, OR, USA, Apr. 1998, pp. 219-239.
- [2] P. Loo and N. G. Kingsbury, "Motion estimation based registration of geometrically distorted images for watermark recovery," in *Proc. SPIE*, vol. 4314, Jan. 2001.
- [3] G. Caner, A. M. Tekalp, G. Sharma, and W. Heinzelman, "Local image registration by adaptive filtering," *IEEE Trans. Image Proc.*, submitted August 2005. Accepted for publication.
- [4] J. J. K. Ó. Ruanaidh and T. Pun, "Rotation, scale and translation invariant spread spectrum digital image watermarking," *Signal Processing*, vol. 66, no. 5, pp. 303-317, May 1998.
- [5] S. Pereira et al., "Template based recovery of fourier-based watermarks using log-polar and log-log maps," in *Proc. IEEE Int. Conf. on Multimedia Comput. and Sys.*, vol. 1, 1999, pp. 870-874.
- [6] C. Y. Lin, M. Wu, J. Bloom, I. Cox, M. Miller, and Y. M. Liu, "Rotation, scale, and translation resilient watermarking for images," *IEEE Trans. Image Proc.*, vol. 10, no. 5, pp. 767-782, May 2001.
- [7] M. Alghoniemy and A. Tewfik, "Image watermarking by moment invariants," in *Proc. IEEE ICIP*, Sept. 2000.
- [8] R. Caldelli, M. Barni, F. Bartolini, and A. Piva, "Geometric-invariant robust watermarking through constellation matching in the frequency domain," in *Proc. IEEE ICIP*, Sept. 2000.
- [9] C. W. Honsinger and S. J. Daly, "Method for detecting rotation and magnification in images," US Pat. #5,835,639, 10 Nov. 1998.
- [10] M. Kutter, "Watermarking resisting to translation, rotation, and scaling," in *Proc. SPIE*, vol. 3528, 1-6 Nov. 1998, pp. 423-431.
- [11] M. Alvarez-Rodriguez and F. Perez-Gonzalez, "Analysis of pilot-based synchronization algorithms for watermarking of still images," *Signal Processing: Image Communication*, vol. 17, pp. 611-633, Sept. 2002.
- [12] D. J. Fleet and D. J. Heeger, "Embedding invisible information in color images," in *Proc. IEEE ICIP*, vol. 1, Oct. 1997, pp. 532-535.
- [13] G. B. Rhoads, "Steganography methods employing embedded calibration data," US Pat. #5,636,292, June 1997.
- [14] S. Pereira and T. Pun, "Robust template matching for affine resistant watermarks," *IEEE Trans. Image Proc.*, vol. 9, pp. 1123-1129, June 2000.
- [15] A. Herrigel, S. Voloshynovskiy, and Y. Rytsar, "The watermark template attack," in *Proc. SPIE*, vol. 4314, Jan. 2001, pp. 394-405.
- [16] M. Kutter, S. K. Bhattacharjee, and T. Ebrahimi, "Towards second generation watermarking schemes," in *Proc. IEEE ICIP*, vol. 1, Oct. 1999, pp. 320-323.
- [17] P. Bas, J.-M. Chassery, and B. Macq, "Geometrically invariant watermarking using feature points," *IEEE Trans. Image Proc.*, vol. 11, no. 9, pp. 1014-1028, Sept. 2002.
- [18] M. U. Celik, E. Saber, G. Sharma, and A. M. Tekalp, "Analysis of feature-based geometry invariant watermarking," in *Proc. SPIE*, vol. 4314, Jan. 2001, pp. 261-268.
- [19] J. Lichtenauer, I. Setyawan, T. Kalker, and R. Lagendijk, "Exhaustive geometrical search and the false positive watermark detection probability," in *Proc. SPIE*, vol. 5020, Jan. 2003, pp. 203-214.
- [20] D. J. Coumou and G. Sharma, "Watermark synchronization for feature-based embedding: Application to speech," submitted to *IEEE ICME*, 2006.
- [21] —, "Insertion, deletion codes with feature-based embedding: A new paradigm for watermark synchronization with applications to speech watermarking," *IEEE Trans. Info. Forensics and Security*, in preparation for submission.
- [22] M. Celik, G. Sharma, and A. M. Tekalp, "Pitch and duration modification for speech watermarking," in *Proc. IEEE ICASSP*, Mar. 2005, pp. II, 17-20.
- [23] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Info. Theory*, vol. 47, no. 4, pp. 1423-1443, May 2001.
- [24] E. Molines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Communication*, pp. 453-467, 1990.
- [25] M. C. Davey and D. J. C. Mackay, "Reliable communication over channels with insertions, deletions, and substitutions," *IEEE Trans. Info. Theory*, vol. 47, no. 2, pp. 687-698, Feb. 2001.
- [26] "Ohio state university speech corpus," Online, accessed Jan 2006. [Online]. Available: <http://buckeyecorpus.osu.edu>
- [27] V. Licks, F. Ourique, R. Jordan, and F. Pérez-González, "The effect of the random jitter attack on the bit error rate performance of spatial domain image watermarking," in *Proc. IEEE ICIP*, vol. 2, Oct. 2003, pp. 455-458.
- [28] M. Barni, "Effectiveness of exhaustive search and template matching against watermark desynchronization," *IEEE Sig. Proc. Ltrs.*, vol. 12, no. 2, pp. 158-161, Feb. 2005.
- [29] N. Merhav, "An information-theoretic view of watermark embedding-detection and geometric attacks," in *WaCha*, June 2005, Barcelona, Spain.
- [30] M. U. Celik, G. Sharma, E. Saber, and A. M. Tekalp, "Hierarchical watermarking for secure image authentication with localization," *IEEE Trans. Image Proc.*, vol. 11, no. 6, pp. 585-595, Jun. 2002.

⁴In practice this step is a soft decoding and outputs probabilistic information rather than a single optimum result from the search.