RC22287 (W0112-039) December 17, 2001
Computer Science

IBM Research Report

Application of Composite Invisible Image Watermarks to Simplify Detection of a Distinct Watermark from a Large Set

Gordon W. Braudaway, Frederick C. Mintzer
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Delhi - Haifa - India - T. J. Watson - Tokyo - Zurich

Application of Composite Invisible Image Watermarks to Simplify Detection of a Distinct Watermark from a Large Set

Gordon W. Braudaway, Fred Mintzer

International Business Machines Corporation, Thomas J. Watson Research Center
Yorktown Heights, New York 10598

ABSTRACT

Earlier, we presented a highly robust invisible watermarking technique for digitized images [1] having a payload of one bit -- indicating the presence or absence of the watermark. Other invisible watermarking techniques also possess this property. This family of techniques may be used to watermark a source image with distinct marks, perhaps to indicate the identity of the recipient, resulting in a set of many *near-copies* of the source image. Then, the problem of detecting a distinct watermark in an image from the set may imply attempting detection of all possible watermarks. In this paper we will present a technique using *composite watermarks* which reduces the number of attempts necessary for distinct watermark detection. If the number of images in the set is *m* to the power *n*, then the number of attempted detections is never more than *m* times *n.* Thus, for *m* = 10 and *n* = 3, a set of 1000 distinctly watermarked near-copies can be produced, but instead of 1000 attempted detection's to insure identification of a particular watermark, only thirty are required. The techniques used for constructing composite watermarks will be detailed and limitations of this approach will be discussed. Results of a successful detection of a distinct watermark from a large set will be presented.

Keywords: image security, invisible watermarking, image processing, fingerprinting

1. INTRODUCTION

In a previous paper [1] we presented a highly robust invisible watermarking technique for digitized images. By design, the watermarking technique has a payload of one binary bit, which indicates the presence or absence of the watermark. We deliberately attempted to sacrifice payload to achieve lower vulnerability to determined attacks aimed at rendering the watermark undetectable. There are other embodiments of invisible watermarking that have this same property. This family of techniques may be used to watermark a source image with distinct watermarks, perhaps to indicate the identity of the recipient, resulting in a set of many *near-copies* of the source image. A problem then arises in detecting which distinct watermark an image from the set might have. In the worst case, it requires attempting detection of all possible watermarks to find a particular one.

In this paper we will present a variation of the image watermarking technique that uses a *composite watermark.* Although composite watermarks are similar in form and have equally strong robustness, their use reduces the number of attempts necessary for distinct watermark detection. When composite watermarks are used to mark all near-copies in a set having *m* to the power *n* near-copies, the number of attempted detections needed to find a particular copy is never more than *m* times *n.* Thus, for *m* = 10 and *n* = 3, a set of 1000 distinctly watermarked near-copies can be produced, but instead of 1000 attempted detection's needed to insure identification of a particular watermark, only thirty are required. This technique can not be extended indefinitely, however. The practical limitation of the value of *n* will be shown to be three to four.

2. A SUMMARY OF THE WATERMARKING TECHNIQUE

A summary of a the referenced robust invisible watermarking technique is presented here as foundation for the following discussion. The watermark that is to be embedded into a digitized image is represented as a rectangular array of numeric elements, called a *watermarking plane*. A watermarking plane has the same number of rows and columns as the digitized source image into which it is to be embedded. The invisible watermark is embedded as a random, but reproducible, small modulation of the luminance of each image pixel; it becomes a permanent part of the watermarked image.

The technique of constructing the watermarking plane is fundamental to insuring the robustness of the embedded watermark and its ability to survive determined attacks. To this end, the procedure by which the values of its elements are chosen must be carefully cast using techniques borrowed from cryptography, mathematical statistics, and two-dimensional signal processing

theory. Pseudo-random values are created from successive pairs of eight-bit groups taken from a cryptographicly secure sequence of bits. The sequence can be reproduced, at will, knowing only the details of its generating method, a specific private cryptographic key and a public seed needed to initialize a particular sequence. The secure sequence is a necessary component of the claim of robustness of the watermark. It is also required to be reproducible so at a future time the reproduced sequence can be used to know what pattern to look for when watermark detection is attempted.

The value of each element in the watermarking plane is a uniformly distributed pseudo-random number determined by linearly mapping successive pairs of eight-bit groups taken from a secure sequence into the domain of 1 to 1-2β, where β is called the *modulation strength.* Element values of the watermarking plane, as an ensemble, are scaled and adjusted to have both a mean and median of 1-β. The watermark is embedded into the source image by multiplying the luminance of each pixel by its corresponding value in the watermarking plane. *Small random luminance modulation without chroma alteration is the essence of invisible watermarking.*

When an image is reduced in size, the high frequency feature content in the source image is diminished in the reduced image. If the applied watermarking plane contains significant high frequency content, that content will be obliterated in the reduced image. Although high frequency content is beneficial in making the watermark less visible, it also makes it vulnerable when a watermarked image is reduced in size. The deliberate suppression of high frequency content in the watermarking plane makes the watermark less vulnerable to typical image manipulation, but it generally makes it more visible by producing a pattern with larger features. The suppression of high frequency content in the watermarking plane can be accomplished by employing techniques derived from two-dimensional signal processing theory. Assured low frequency feature content with a maximum frequency only one fourth that possible in an image plane of its size is considered adequate. The reader is referred to the referenced paper [1] for further detail on how this can be accomplished.

3. A SUMMARY OF WATERMARK DETECTION

Detection of an embedded watermark is a daunting task, especially after manipulation of the watermarked image. It requires detecting the presence of a particular known small modulation of a random two-dimensional carrier, where the carrier is composed of the pixel luminance values of the unmarked source image.

The first challenge in detecting a watermark is to determine in what ways a watermarked image may have been manipulated. It may have been cropped, reduced nonlinearly in size, and rotated through a small angle. A previous paper [2] details how a digitized image, suspected of being a watermarked image from the large set, can be realigned automatically by referencing its source image. Since the source image and each watermarking plane were geometrically aligned with each other at the time of embedding, the realigned watermarked image will also be aligned with an expected watermarking plane, and detection of the expected watermark can be begun.

The process of watermark detection is designed to establish a statistical probability, which approaches certainty, that an expected watermark was, in fact, embedded into the image. This is implemented by examining the luminance of every pixel in the watermarked image and comparing it to the average luminance of its neighboring pixels. A useful neighborhood is an 11 by 11 pixel square, so each pixel has 120 neighbors. The corresponding element from the reconstructed watermarking plane is also compared to the average of its neighboring elements in a neighborhood of the same size. The statistical correlation of the suspected watermarked image with a candidate watermarking plane is established by coincidence counting. If the luminance of a pixel is greater than the average luminance of its neighboring pixels, and the value of the corresponding element from the watermarking plane is greater than the average value of its neighboring elements, the agreement is considered to be a coincidence and a single coincidence counter, initially set to zero, is incremented by one. Correspondingly, if both are less than their neighborhood averages, the agreement is also considered to be a coincidence and the coincidence counter is also incremented by one.

But if the luminance of a pixel is greater than the average luminance of its neighboring pixels, and the value of the corresponding element from the watermarking plane is less than or equal to the average value of its neighboring elements, or vise-versa, the disagreement is considered to be a non-coincidence and the coincidence counter is decremented by one. Thus, if a watermarking plane is embedded into an image having uniform finite pixel values, it should now be apparent, because of the careful statistical construction of the watermarking plane, that the expected value of the coincidence count should be equal to the number of pixels in the image, and the expected value of the coincidence count should be zero if no watermarking plane has been embedded. This can be verified by embedding a watermark into an image having uniform gray pixel values and attempting detection by the technique described.

Using source images derived from natural scenes, which can have highly varying pixel values, the count in the coincidence counter is not as distinct. The inherent variability of the pixel values in the small neighborhood regions are a significant source of noise, and cause the coincidence counter to have values other than the theoretically expected values. But fortunately for images of interest (other than artificial images possessing pure noise for pixel values) there is enough relative constancy in enough small neighborhoods to usually allow unequivocal detection of an embedded watermark with a mathematical probability that approaches certainty (a large positive count), and an unequivocal non-detection (a very small count, either positive or negative) if the expected watermark is not embedded into the image.

For visual verification of watermark detection, an array of counters is used instead of a single counter. For example, an array of counters having 32 rows and 128 columns (4096 counters in total) might be used for images having at least one million pixels. Counters are selected in a fixed random sequence, and the evaluation of each sequential pixel along with its corresponding element in the watermarking, as described above, causes the next counter in the fixed random sequence to be incremented or decremented. Clearly, the count expected in each of these counters for a watermark detection is 1/4096 of the number of pixels in the image, and in natural images, the counter values will necessarily be more vulnerable to the noise inherent in the image pixels than the count in a single counter. But as an ensemble, the array of counters conveys the same information as the single counter. (The algebraic sum of the 4096 counter values would be equal to the count in a single counter).

To assist in the visual judgment of detection or non detection, three binary images are formed. The first binary image, called the *coincidence image*, is formed having the same dimensions as the counter array (e.g., 32x128 pixels). The algebraic sign of each coincidence counter value is used to determine a corresponding binary pixel value in the coincidence image. The algebraic sign is mapped to a white pixel value if it is positive and a black value if it is negative. Although the coincidence image contains all the necessary information needed to make a detection judgment, its appearance as a “salt and pepper” scatter pattern that is not particularly useful. The ability of the human visual system to recognize a pattern in a scatter diagram can be exploited to assist in this judgment. To do this, a second binary image, called a *visualizer*, is formed having the same dimensions as the coincidence image. A bold and clearly recognizable binary pattern is placed into the visualizer. A typical visualizer image is shown in Figure 1.

Figure 1. A Typical Visualizer

Pixel values in the coincidence image are then combined, pixel by pixel, with corresponding values in the visualizer image to produce pixels of the third image, called the *visualizer-coincidence* image. The combining operation is an exclusive-or. By this it is meant that for every black pixel in the coincidence image, the corresponding visualizer pixel is inverted, white to black or black to white, and placed in the corresponding pixel location in the visualizer-coincidence image. For every white pixel in the coincidence image, the corresponding visualizer pixel is copied into the corresponding pixel location in the visualizer-coincidence image without alteration. A judgment is then made as to whether the pattern previously placed in the visualizer is recognizable in the visualizer-coincidence image.

In a watermarked image, even a highly textured one, the visualizer’s binary pattern is almost always clearly recognizable in the visualizer-coincidence image if the expected watermark is present. For a highly textured test image, a typical coincidence images is shown in Figure 2, and the visualizer-coincidence image produced by the process just described is shown in Figure 3. Visual recognition of the pattern in the visualizer-coincidence images signifies a highly credible detection of the presence of a known watermark in the image.

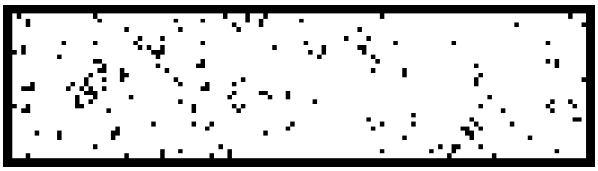

Figure 2. A Typical Coincidence Image ($\beta = 2.5$ %).

Figure 3. A Typical Visualizer-Coincidence Image ($\beta = 2.5$%).

Increasing modulation strength is an effective means of making an embedded watermark more readily detectable. An unfortunate side effect of increasing modulation strength, however, is that it tends to make the embedded watermark visible.

An attempt to detect the presence of a watermark in an image not having one, or in an image having one but for which the watermarking plane can not be correctly reconstructed, produces a pattern in the visualizer-coincidence image that is an unrecognizable random melee. A visualizer-coincidence images produced from an unmarked image is shown in Figure 4. An attempt to detect a watermark using a watermarking plane reconstructed with an incorrect cryptographic key produced a visualizer-coincidence image that is shown in Figure 5.

Figure 4. Visualizer-Coincidence Image from an unmarked image.

Figure 5. Visualizer-Coincidence Image using an incorrect key.

4. CONSTRUCTING A COMPOSITE WATERMARK

Watermarks are embedded into an image by multiplying the luminance of the components of each pixel of the image by a corresponding value of a watermarking plane, w(i,j), where $1 > w(i,j) \geq (1-2\beta)$, i is the value's row index, j is the value's column index, and β is the modulation strength of the watermark. More than one watermark can be embedded into an image by sequential applications of the technique presented above. Each watermark so embedded will be represented by a distinct watermarking plane generated using its own distinct method, key and seed. If the quantity **P** of watermarking planes, designated $w_p(i,j)$, each having a modulation strength β_p, for p=1, 2,… ,**P**, are sequentially embedded into an image, then the presence of each watermarking plane can be detected by knowing only its distinct method, key and seed from which its distinct watermarking plane can be reconstructed. The presence of other watermarks embedded into the image produces only a small degree of additional random noise in the detection process because each of the **P** watermarking planes is statistically uncorrelated with each other.

Using the mathematical property of association, a single *composite watermarking plane*, $w_C(i,j)$, can be constructed as the element-by-element product

$$w_C(i,j) = w_1(i,j) v w_2(i,j) v \ldots v w_P(i,j)$$

To a first order of approximation, the equivalent modulation strength of the *composite watermarking plane* is the sum of the modulation strengths of its parts, or

$$\beta_C = \beta_1 + \beta_2 + \ldots + \beta_P$$

Applying several distinct watermarks to an image by using a composite watermarking plane can be very useful. For example, if thirty distinct watermarking planes are generated, and the thirty are divided into three equal groups in such a manner that each watermarking plane appears in only one group, then 1000 distinct composite watermarking planes can be produced by using one watermarking plane from each of the three groups. If the 1000 composite watermarking planes are used to watermark 1000 copies of an image, subsequent detection of the single composite watermark requires no more than thirty detection

attempts using each of the thirty distinct watermarking planes. Since each distinct watermarking plane, or composite part, of a composite watermarking plane is individually detectable, a particular composite watermarking plane can also be identified by detecting its three composite parts. Once the three composite parts are individually detected, a final (all be it unnecessary) confirming detection of the particular composite watermarking may be attempted as further verification of the correct identification of the one of 1000 distinct composite watermarking planes. This is significantly more efficient than searching for the particular watermark by attempting detection of each of the 1000 composite watermarking panes until the correct one is found. On average, detection of a particular composite watermarking plane will require fifteen tries rather than 500, and a maximum of thirty tries rather than 1000. Applying more than one distinct watermark to an image is also useful when the seed and key of one watermark remain secret to a first party while, for example, a second seed and key are divulged to second party. The divulged seed and key could be used by a near-copy recipient to detect the second watermark, but the secret watermark remains undetectable except by the first party having knowledge of the secret seed and key used to generate it.

5. AN EXAMPLE APPLICATION OF THE COMPOSITE WATERMARK TO A LARGE SET

As an example application, thirty distinct watermarking planes are generated using the technique described in Reference 1. For this discussion, the method of generating the secure random sequence for all watermarks will be assumed the same, although in actual practice this need not be the case. Each of the thirty watermarking planes, designated as $w_p(i,j)$, p = 1, 2, …, 30, is generated using its own distinct key and seed. A particular key and seed will be referred to hereon as a *parameter pair*. As stated before, each watermarking plane can be reconstructed at any future time if its distinct parameter pair is then known. The thirty distinct watermarking planes are divided arbitrarily into three equal groups with none of the thirty appearing in more than one group. Members of the three groups are designated as $w_q(i,j)$, $w_r(i,j)$, and $w_s(i,j)$, respectively. Once divided, knowledge of the specific parameter pair corresponding to each member of each group is retained, since regeneration of the thirty watermarking planes, each from its specific parameter pair, will be required for watermark detection at a later time. Using the three groups, 1000 distinct composite watermarking planes, designated $w_C(i,j)$, are formed as the element-by-element products of three distinct watermarking planes, one chosen from each of the three groups.

The 1000 composite watermarking planes are the maximum number of distinct combinations of members that can be formed from the three groups in the manner specified. It is apparent that 10,000 distinct composite watermarking planes could be formed in a similar manner from forty distinct watermarking planes separated into four equal groups. The 1000 distinct composite watermarking planes can be used one at a time to embed a distinct watermark into 1000 copies of a single image, using the technique described in Reference 1.

At a future time, if it is desired to determine whether one of the 1000 composite watermarks was used to embed a watermark into an image suspected of being one of the 1000 near-copies, knowledge of the three groups of parameter pairs saved during the watermarking process is recalled. The technique used for watermark detection is described in summary above and in detail in Reference 1. It is not necessary to attempt watermark detection with each of the 1000 composite watermarking planes until the correct one is found. Rather, the component parts of the composite watermarking plane, namely the three watermarking planes used in its construction, can be used individually instead. This limits the search to not more than thirty tries, each try using a different one of the watermarking planes from the three groups.

A color source image, having 2184 columns and 2277 rows of 24-bit pixels, was watermarked using a composite watermarking plane constructed from three component watermarking planes. It is representative of one of the possible 1000 watermarked near-copies of the source image. The three component watermarking planes were taken from three groups of watermarking planes designated A, B, and C, each group having ten members numbered 1 through 10, and were, specifically, [Group A, Member 3], [Group B, Member 8], and [Group C, member 5] . The composite modulation strength, β_C, was 3%.

The process of detection is begun by choosing a previously unused parameter pair from the first group of ten parameter pairs. A unique watermarking plane is reconstructed from the chosen parameter pair. Detection of the reconstructed watermarking plane in the suspected near-copy is attempted, and if the detection is successful, the index of the parameter pair, q^*, from the first group of parameter pairs that was used to regenerate the watermarking plane is saved and no further selection is made from the first group of parameter pairs; otherwise a test is made to determine if all ten of the first group of parameter pairs have been used. If not, then another parameter pair is chosen from the first group and another detection is attempted. This continues until either a detection is made or all ten parameter pairs have been used. If no detection is made and all ten parameter pairs have been tried, q^* is set to “not found” to reflect no detection from the first group of parameter pairs.

In a similar manner, the process continues with selection of a parameter pair from the second group of ten parameter pairs and *r** is set to the index of the parameter pair where detection occurs or to "not found" to reflect a detection or no detection from the second group of parameter pairs. The process is further continued by choosing a previously unused parameter pair from the third group of ten parameter pairs and *s** is set to the index of the parameter pair where detection occurs or to "not found" to reflect a detection or no detection from the third group of parameter pairs.

If the suspect near-copy is in fact watermarked, on average only five choices from each group of parameter pairs will be needed for a total of fifteen attempted detections, and not more than thirty detections will ever be needed. The indexes of the detected watermark from each of the three groups are designated *q**, *r** and *s**, and the three detected watermarking planes are $w_{q^*}(i,j)$, $w_{r^*}(i,j)$, and $w_{s^*}(i,j)$. If *q**, *r** and *s** exist, that is, if none of them was set to "not found", the distinct composite watermarking plane is determined, and is $w_{C^*}(i,j) = w_{q^*}(i,j) \vee w_{r^*}(i,j) \vee w_{s^*}(i,j)$. A further confirming verification can be done by detecting the composite watermarking plane $w_{C^*}(i,j)$, but, strictly speaking, it is not necessary. There can be only one composite watermarking plane composed of the detected component parts designated by $w_{q^*}(i,j)$, $w_{r^*}(i,j)$, and $w_{s^*}(i,j)$.

Referring to Figure 6, and using the watermark detection technique described in Reference 1, the visualizer-coincidence images resulting from attempts to detect thirty distinct watermarks in a near-copy are separated into three groups of ten, represented by the three columns. The positions of the visualizer-coincidence images in these three groups are related one-to-one to the positions of the distinct watermarking planes in their three groups. Hence, the clearly identifiable patterns in the visualizer-coincidence images of [Group A, Member 3], [Group B, Member 8], and [Group C, Member 5] identify the successfully detected component parts of a composite watermarking plane. The remaining twenty-seven visualizer-coincidence images result from attempted detections that were unsuccessful. The final confirming visualizer-coincidence image, the lone image at the bottom of the Figure 6, is that obtained using the composite watermark reconstructed from its three detected component parts. It is a slightly more positive detection.

In the present example, detection of only one of three distinct embedded watermarks is strong evidence of knowledge of its parameter pair, but 100 of the 1000 near-copies would have that embedded watermark and isolation to a single near-copy is not possible by this means. Detection of only two of the three distinct embedded watermarks is strong evidence of knowledge of the two parameter pairs, but ten of the 1000 near-copies would have those embedded watermarks and isolation to the single near copy is again not possible. However, detection of all three distinct embedded watermarks is strong evidence of knowledge of all three parameter pairs, and isolation to the distinct near-copy from the 1000 near-copies is achieved. Detection of three distinct embedded watermarks, one from each group, is logically equivalent to detecting the single composite watermark that is constructed using those three distinct watermarks as its component parts.

It should be noted that there are detection strategies using component parts of a composite watermark other than the strategy shown in the present example. Some of those strategies may make more effective use of the component parts to enlarge the group of composite watermarks that can be formed; but, in so doing, they may also incur the cost of a greater average number of attempted detections needed to find an embedded composite watermark from that larger group. For example, if there are twenty distinct component parts, the maximum number of combinations of twenty component parts taken three at a time is 1140, and if there are thirty, that maximum number rises to 4060.

It should also be noted that the composite watermark embedding and detection technique is usable with other watermark embedding and detection techniques by sequentially applying the logical equivalents of their watermarking planes, whether those equivalents lie in the spatial pixel domain or in a transform domain (for example, a Discrete Fourier transform domain or a Discrete Cosine transform domain). To this end an alternative technique for embedding a single watermark or several watermarks into a digital image can be used for watermarking techniques that are substantially different from the one specified in Reference 1. One alternative technique is accomplished by first defining an auxiliary monochrome image plane. The size of the auxiliary monochrome image plane is chosen such that pixels in the monochrome image plane have a one-to-one correspondence with pixels of the source image. The luminance values of pixels in the monochrome image are chosen initially to be uniform and greater than zero. A first watermark is embedded into the auxiliary monochrome image plane forming a first "watermarking" image. Another watermark can be embedded by sequentially embedding a watermark into the first watermarking image forming another watermarking image. This can be repeated a third time (and possibly a fourth) if desired forming a final watermarking image. The luminance pixel values of the final watermarking image are linearly mapped such that all luminance values v(i,j) lie in a domain $1 > v(i,j) \geq (1-2\beta_C)$, where i is the value's row index, j is the value's column

index, and β_C is the equivalent cumulative modulation strength of the one or several watermarks. It should be apparent that after the pixel values of the final watermarking image are linearly mapped, v(i,j) is equivalent in form, meaning, and properties to the *composite watermarking plane*, $w_C(i,j)$, described previously, and v(i,j) can therefore serve as an alternative form of a composite watermarking plane. It should also be apparent that almost any image watermarking technique can be used to embed watermarks sequentially into the auxiliary monochrome image plane to form a final watermarking image that is subsequently used to form the linearly mapped values v(i,j).

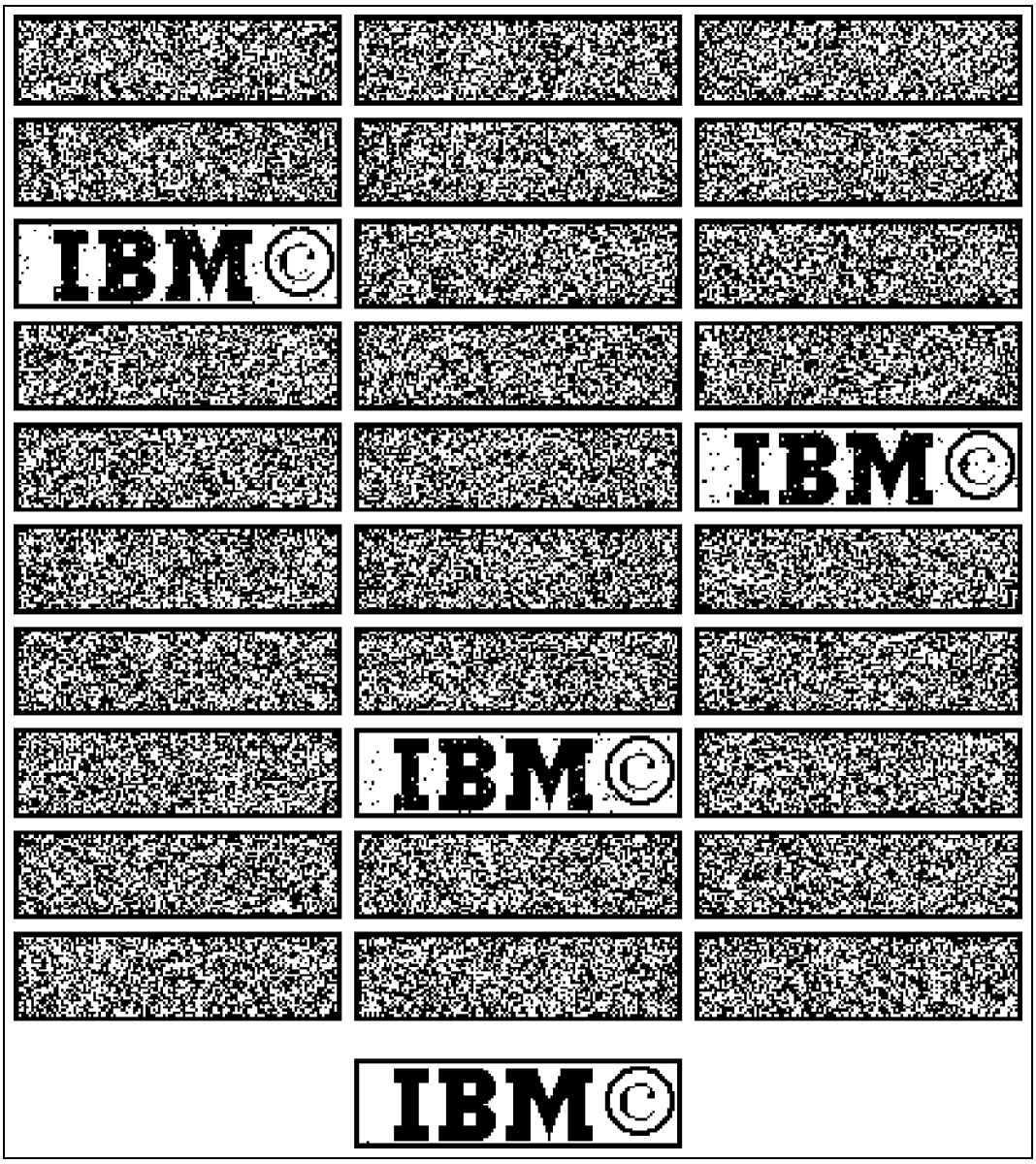

Figure 6. Three groups of ten visualizer-coincidence images from watermark detections attempted to isolate one watermark from 1000 possible combinations. It shows detection of the component parts of a Composite Watermark generated using key and seed parameter pairs from [Group 1, Member 3], [Group 2, Member 8] and [Group 3 Member 5]. The lone visualizer-coincidence image at the bottom of this figure is that of the confirming composite watermark detection