# Compression Tolerant Image Authentication

*Sushil Bhattacharjee* *Martin Kutter*
Signal Processing Laboratory
Swiss Federal Institute of Technology
1015 Lausanne, Switzerland
{sushil.bhattacharjee, martin.kutter}@epfl.ch

**Abstract**

*It is straightforward to apply general schemes for authenticating digital data to the problem of authenticating digital images. However, such a scheme would not authenticate images that have undergone lossy compression, even though they may not have been manipulated otherwise. In this paper we propose a scheme for authenticating the* visual content *of digital images. This scheme is robust to compression noise, but will detect deliberate manipulation of the image-data. The proposed scheme is based on the extraction of feature-points from the image. These feature-points are defined so as to be relatively unaffected by lossy compression. The set of feature-points from a given image is encrypted using public key encryption, to generate the digital signature of the image. Authenticity is verified by comparing the feature-points of the image in question, with those recovered from the previously computed digital signature.*

## 1 Introduction

Verification of data integrity has become a very important issue in the digital age. This is due to the fact that one may easily modify digital data without leaving tracks. Cryptographic schemes have been developed to authenticate the content of digital data. These schemes, also called digital signatures, are usually based on hashing (for data reduction), followed by public key encryption. These authentication schemes are applicable to any kind of digital data and have proven to be secure.

These general digital authentication schemes can, in principle, also be used to authenticate digital images. However such an image authentication scheme cannot accommodate lossy image compression. Although lossy image compression does not alter the image authenticity, in the sense of its visual content, authentication with any scheme involving hashing will fail. For visual data authentication, hashing is not suitable since it enforces authenticity of individual bits, not of the visual content. Lossy compression methods retain the authenticity of the visual content of an image but do not guarantee color-authenticity of individual pixels. An image authentication scheme should be tolerant to compression noise, but should detect modifications of the visual image content. We consider such a system to be *compression tolerant.*

Previous work on compression tolerant image authentication has been reported by Schneider et al. [4]. They have proposed a scheme based on image-block histograms to authenticate the visual content in an image. In the authentication process the Euclidean distance between histogram of each block in the original image and the histogram of the corresponding block in the image under inspection is computed. The sum of all such distances over the entire image is used as a measure of image authenticity. If this sum is less than a pre-specified threshold, the image in question is considered authentic. This approach requires the storage of the public key encrypted histograms, which may be considerably large. The most significant drawback of their method is that it is not very secure since it is trivial to modify an image without altering its histogram.

In this paper we propose a technique for authenticating digital images that uses features inherent to the image. The process of extracting the image-features to be used for authentication is fairly robust to compression noise. Therefore, the feature values do not change even if the image undergoes a compression/decompression cycle.

In Section 2 we briefly describe the proposed system for image authentication. The process of extracting the image-features used for authentication is described in Section 3, and the way in which these features are used for determining the authenticity of a given image is described in Section 4. In Section 5 we present results of some experiments to demonstrate the viability of the proposed method.

## 2 Proposed Authentication System

The process of image authentication consists of two parts: generation of the *digital signature* of the original image, and subsequent verification of the digital signature of the image to be authenticated with the stored signature of the original. The two parts are briefly discussed below.

### 2.1 Digital Signature Generation

As in other cryptographic authentication schemes, our scheme for generating the digital signature of an image involves two steps. First, the coordinates of visually important image structures are extracted. These coordinates give the feature-points on which the authentication scheme is based. The procedure for extracting the feature-points of an image is described in Section 3. The feature extraction can be compared to the hashing procedure in general authentication systems, in the sense that it is a many-to-one mapping of the image-content that results in data reduction. In the second step the set of feature-points is encrypted with a public key encryption scheme such as RSA [5] using a private encryption key. The result, which is the digital signature of the image, can then be stored in a database (or added to the image header) for future use in verification.

### 2.2 Verification

In the verification process, we test the truth of the assertion that a given image, $A$, is an un-manipulated version of some other, known image, $B$. To verify the authenticity of image $A$, first the feature-points of $A$ are computed. The digital signature of the image $B$ is decrypted to obtain the feature-points of $B$. The two sets of feature-points are compared to verify the authenticity of $A$. If the features match, image $B$ is considered to be authentic. However, if some features do not match the image has been modified and is not authentic anymore.

## 3 Feature Extraction

The image authentication scheme proposed here relies on stable image features extracted from the original image. The feature extraction technique is inspired by the work of Manjunath and colleagues [3, 6]. They attempted to extract "perceptually interesting" image features which were then used in applications such as image registration, motion correspondence, and even human-face recognition. Their feature-detection scheme is based on the so-called *scale interaction model*. The scale interaction model described in [3] utilizes Gabor wavelets at two different scales (or resolutions). For the sake of completeness, a brief overview of the scale interaction model follows.

### 3.1 Scale Interaction Model

Given an image, $A$, let $W_i(\vec{x},\theta)$ and $W_j(\vec{x},\theta)$ represent the Gabor wavelet transform coefficients at image location $\vec{x}$ in the preferred orientation $\theta$, for two scales, $i$ and $j$ respectively. The feature-detection function, $Q_{ij}()$, is then defined as

$$Q_{ij}(\vec{x},\theta) = f(W_i(\vec{x},\theta) - \gamma W_j(\vec{x},\theta)),$$

where $f(\cdot)$ represents some nonlinear transformation function. $\gamma$ is a normalizing factor which depends on the difference between the two scales $i$ and $j$. Manjunath et al. [3] then consider an image location $(\vec{x})$ as a potential feature-point if the function $Q_{ij}()$ has a local maxima at the point $(\vec{x})$. In mathematical terms, a feature-point at $(\vec{x})$ should satisfy the following criterion:

$$Q_{ij}(\vec{x},\theta) = \max_{(\vec{x}')\in N_{\vec{x}}} Q_{ij}(\vec{x}',\theta),$$

where $N_{\vec{x}}$ represents the local neighborhood of the point $(\vec{x})$. The scale-interaction based procedure for identifying image-feature locations is related to the end-stopping behavior of hyper-complex cells in the visual cortex[2, 3].

### 3.2 Proposed Scheme for Feature Extraction

We use a similar approach for identifying image features for our purposes. The main difference between the approach used by Manjunath et al. and our approach is that we use Mexican-Hat wavelets instead of Gabor wavelets. In two dimensions, the response of the Mexican-Hat mother wavelet, $\psi(\vec{x})$, is defined as:

$$\psi(\vec{x}) = (2 - |\vec{x}|^2) exp(-\frac{\vec{x}^2}{2}).$$

It is the negative Laplacian of a Gaussian, and is also referred to as the Marr wavelet. The isotropic nature of the Mexican-Hat filter is better suited for detecting point-features than the Gabor filter[1]. Using the Gabor filter based approach described above, a potential feature point $(\vec{x})$ is selected as an image feature if it satisfies the following criterion[6]:

$$\max_{\theta} Q_{ij}(\vec{x},\theta) > T,$$

where $T$ is a user-defined threshold. In words, the above criterion considers several orientations, and accepts a location as feature point if the maximum in any direction is larger than a threshold. The Mexican-Hat filter inherently responds to point-features, and is not direction sensitive. Therefore the resulting feature-detection scheme is less complex than the scale-interaction model discussed in Section 3.1. In

fact, for our purpose, we have found the Mexican-Hat based scheme to produce more stable results than those produced by the equivalent Gabor filter based scheme.

Our approach for detecting feature-points is as follows.

1. Define the feature-detection function, $P_{ij}()$ as:

$$P_{ij}(\vec{x}) = |M_i(\vec{x}) - \gamma \cdot M_j(\vec{x})|,$$

where $M_i(\vec{x})$ and $M_j(\vec{x})$ represent the responses of Mexican-Hat wavelets at the image-location $(\vec{x})$ for scales $i$, and $j$ respectively. For an image $A$, the wavelet response $M_i(\vec{x})$ is given by:

$$M_i(\vec{x}) = \langle (2^{-i}\psi(2^{-i} \cdot \vec{x})); A) \rangle,$$

where $\langle \cdot ; \cdot \rangle$ denotes the convolution of its operands. We only consider wavelets on a dyadic scale. Thus, the normalizing constant, $\gamma$, is given by $\gamma = 2^{-(i-j)}$. The operator $|\cdot|$ returns the absolute value of its parameter.

For the experiments reported in this paper, we have used $i = 2$ and $j = 4$. The choice of these values was influenced by the following considerations:

- the scales used should be as far apart as possible, for good results;
- the coarsest possible scale ($i = 1$) is not suitable as it is too sensitive to global variations;
- the finer the scale the more sensitive it is to distortions such as quantization noise. Thus, in order to make the scheme robust to moderately lossy compression (good quality JPEG), the finer scale, $j$, should not be too fine.

2. Determine points of local maxima of $P_{ij}()$. These maxima correspond to the set of potential feature-points. For the results described in this paper, we have used a circular neighborhood with a radius of 5 pixels, to determine the local maxima of $P_{ij}()$.

3. Accept a point of local maximum in $P_{ij}()$ as a feature-point if the variance of the image-pixels in the neighborhood of the point is higher than a threshold. This criterion helps avoid spurious local maxima in featureless regions of the image. We have used a $7 \times 7$ neighborhood around the point, for computing the local variance. In this work, a candidate point is accepted as a feature-point if the corresponding local variance is less than 10.

the resulting feature-points are ordered by row and column poistion. A string of digits is constructed by concatenating the column- and row-positions of these points. This series is encrypted using private-key encryption, to generate the image signature (see Section 2).

## 4 Feature Comparison

In order to determine whether an image $A$ is authentic with respect to another known image, $B$, the following procedure is used. The set, $S_A$, of feature-points is computed from $A$. This is compared with the set of points, $S_B$, obtained by decrypting the digital signature of $B$. To authenticate the image we verify that each feature-location present in $S_B$ is also present in $S_A$ and further that no feature-location present in $S_A$ is absent in $S_B$. While comparing locations of feature-points, we tolerate a deviation of up to 1 pixel in order to make the system less sensitive to compression noise. Two feature-points with coordinates $\vec{x}$ and $\vec{y}$ are said to match if

$$|\vec{x} - \vec{y}| \ < \ 2.$$

## 5 Experimental Results

In this section we show results of the proposed scheme on a sample image. The image-authentication scheme should be able to detect any editing applied to the original image, but should be robust to lossy compression. Figure 6 shows the feature-points extracted for different versions of the test image. The original image is shown in Figure 6(a), and the corresponding set of feature-points is shown in Figure 6(b). Figure 6(c), (d), and (e) show the effect of adding a logo on the top-right corner of the original image. The set of feature-points corresponding to modified image of Figure 6(c) is shown in Figure 6(d). The difference between Figure 6(d) and Figure 6(b) is shown in Figure 6(e). Note that the points shown in Figure 6(e) all correspond to the edited region of the image. The points in Figure 6(e) do not quite coincide with the position of the logo. This is due to the relatively large-scale wavelets used in this work. The two points appearing at the bottom-right corner are also a result of the logo on the top-right. This wrap-around effect occurs, again, due to the large scale of the analyzing wavelets, and the fact that the Discrete Fourier Transform makes implicit assumptions about the periodicity of the input signal. (The procedure for computing the function $P_{ij}()$ is implemented in frequency domain.) Figure 6(f), (g), and (h) show the result when the original image is edited to remove image structures. Figure 6(f) is a version of the original image where

the rock in the center of the image has been removed. Figure 6(g) shows the set of feature-points corresponding to Figure 6(f) and Figure 6(h) shows the difference between Figure 6(g) and Figure 6(b). Again, the points shown in Figure 6(h) are all localized to the edited region of the image, but do not correspond to the exact location of the missing rock, because of the size of the analyzing wavelet. It is important to note, however, that the modification does affect the feature-points of the image, even though rock in the original image does not have a really strong contrast with the background.

Even if lossy JPEG compression is applied to the original image such that the decompressed image does not contain visible artifacts introduced by the compression, the proposed method is able to authenticate the image. Figure 6(i) shows the JPEG compressed version of the original image at a JPEG quality of 80%. The corresponding set of feature-points is shown in Figure 6(j). All the points are within 1 pixel of the corresponding points in Figure 6(b).

## 6 Conclusions

In this paper we have proposed a new image authentication scheme that is tolerant to lossy image compression but can detect image editing operations. This scheme relies on visually salient image features which are extracted using scale interaction based on Mexican-Hat wavelets. The results so far are very encouraging. In the experiments reported in this paper, the proposed technique is able to detect the removal of the rock, which is not a region of very strong contrast.

So far, we have experimented with gray-scale images only. Extension to color images, and even video sequences, is straightforward, and involves repeating the exercise for every color-band. Ongoing experiments are aimed at establishing the limits of the proposed authentication scheme. The proposed approach can also be extended to accomodate scaling and other such transformations, which are do not really attack the image-content.

## References


[1] J.-P. Antoine, P. Vandergheynst, and R. Murenzi. Two-dimensional directional wavelets in image processing. *Int. J. Imaging Sys. and Tech.*, 7:152–165, 1996.

[2] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J. Neurophysiology*, 28:229–289, 1965.

[3] B. S. Manjunath, C. Shekhar, and R. Chellappa. A new approach to image feature detection with applications. *Pattern Recognition*, 29(4):627–640, 1996.

[4] M. Schneider and S-F. Chang. A robust content based digital signature for image authentication. In *Proc. ICIP-96*, volume 3, page 227, September 1996.

[5] Douglas R. Stinson. *Cryptography.* CRC Press, 1995.

[6] Q. Zheng and R. Chellappa. A computational vision approach to image registration. *IEEE Trans. Image Processing*, 2(6):311–326, 1993.


Figure 1: Authentication of "Piano" image: **(a)** original image; **(b)** feature-points of (a); **(c)** original image modified on the top right corner; **(d)** feature-points of (c); **(e)** difference between (d) and (b); **(f)** original image edited to remove image features near the center of the image; **(g)** feature points of (f); **(h)** difference between (g) and (b); **(i)** JPEG compressed/decompressed version of (a) with 80% quality; **(j)** feature-points of (i). Note that the points in (e) and (h) correspond to only the image-regions modified in (c) and (f) respectively. The points in (j) are within 1 pixel of the corresponding points in (b).