9-1-2004

Creating and Detecting Doctored and Virtual Images: Implications to The Child Pornography Prevention Act

Hany Farid
Dartmouth College

Creating and Detecting Doctored and Virtual Images: Implications to The Child Pornography Prevention Act

Hany Farid
Department of Computer Science and
Center for Cognitive Neuroscience
Dartmouth College
Hanover NH 03755

Abstract

The 1996 Child Pornography Prevention Act (CPPA) extended the existing federal criminal laws against child pornography to include certain types of "virtual porn". In 2002, the United States Supreme Court found that portions of the CPPA, being overly broad and restrictive, violated First Amendment rights. The Court ruled that images containing an actual minor or portions of a minor are not protected, while computer generated images depicting a fictitious "computer generated" minor are constitutionally protected. In this report I outline various forms of digital tampering, placing them in the context of this recent ruling. I also review computational techniques for detecting doctored and virtual (computer generated) images.

Hany Farid, 6211 Sudikoff Lab, Computer Science Department, Dartmouth College, Hanover, NH 03755 USA (email: farid@cs.dartmouth.edu; tel/fax: 603.646.2761/603.646.1672). This work was supported by an Alfred P. Sloan Fellowship, a National Science Foundation CAREER Award (IIS-99-83806), a departmental National Science Foundation Infrastructure Grant (EIA-98-02068), and under Award No. 2000-DT-CS-K001 from the Office for Domestic Preparedness, U.S. Department of Homeland Security (points of view in this document are those of the author and do not necessarily represent the official position of the U.S. Department of Homeland Security).

Contents

1 Introduction

The past decade has seen remarkable growth in our ability to capture, manipulate, and distribute digital images. The average user today has access to high-performance computers, high-resolution digital cameras, and sophisticated photo-editing and computer graphics software. And while this technology has led to many exciting advances in art and science, it has also led to some complicated legal issues. In 1996, for example, the Child Pornography Prevention Act (CPPA) extended the existing federal criminal laws against child pornography to include certain types of "virtual porn". In 2002 the United States Supreme Court found that portions of the CPPA, being overly broad and restrictive, violated First Amendment rights. The Court ruled that images containing an actual minor or portions of a minor are not protected, while "computer generated" depicting a fictitious minor are constitutionally protected. This ruling naturally leads to some important and complex technological questions – given an image how can we determine if it is authentic, has been tampered with, or is computer generated?

In this report I outline various forms of digital tampering, and review computational techniques for detecting digitally doctored and virtual (computer generated) images. I also describe more recent and emerging technologies that may further complicate the legal issues surrounding digital images and video.

2 Digital Tampering

It probably wasn't long after Nicéphore Niépce created the first permanent photographic image in 1826 that tampering with photographs began. Some of the most notorious examples of early photographic tampering were instigated by Lenin, when he had "enemies of the people" removed from photographs, Figure 1. This type of photographic tampering required a high degree of technical expertise and specialized equipment. Such tampering is, of course, much easier today. Due to the inherent malleablilty of digital images, the advent of low-cost and high-performance computers, high-resolution digital cameras, and sophisticated photo-editing and computer graphics software, the average user today can create, manipulate and alter digital images with relative ease. There are many different ways in which digital images can be manipulated or altered. I describe below six different categories of digital tampering – the distinction between these will be important to the subsequent discussion of the U.S. Supreme Court's ruling on the CPPA.

Figure 1: Lenin and Trotsky (top) and the result of photographic tampering (bottom) that removed, among others, Trotsky.

Figure 2: An original image (top) and a composited image (bottom). The original images were downloaded from `freefoto.com`.

2.1 Composited

Compositing is perhaps the most common form of digital tampering, a typical example of which is shown in Figure 2. Shown in the top panel of this figure is an original image, and shown below is a doctored image. In this example, the tampering consisted of overlaying the head of another person (taken from an image not shown here), onto the shoulders of the original kayaker.

Beginning with the original image to be altered, this type of compositing was a fairly simple matter of: (1) finding a second image containing an appropriately posed head; (2) overlaying the new head onto the original image; (3) removing any background pixels around the new head; and (4) re-touching the pixels between the head and shoulders to create a seamless match. These manipulations were performed in Adobe Photoshop, and took approximately 30 minutes to complete. The credibility of such a forgery will depend on how well the image components are matched in terms of size, pose, color, quality, and lighting. Given a well matched pair of images, compositing, in the hands of an experienced user, is fairly straight-forward.

2.2 Morphed

Image morphing is a digital technique that gradually transforms one image into another image. Shown in Figure 3, for example, is the image of a person (the source image) being morphed into the image of an alien doll (the target image). As shown, the shape and appearance of the source slowly takes on the shape and appearance of the target, creating intermediate images that are "part human, part alien". This morphed sequence is automatically generated once a user establishes a correspondence between similar features in the source and target images (top panel of Figure 3). Image morphing software is commercially and freely available – the software (xmorph) and images used in creating Figure 3 are available at `xmorph.sourceforge.net`. It typically takes approximately 20 minutes to create the required feature correspondence, although several iterations may be needed to find the feature correspondence that yields the visually desired morphing effect.

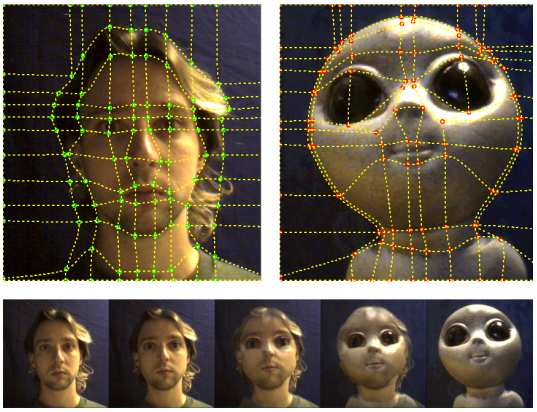

Figure 3: Shown on top are two original images overlayed with the feature correspondence required for morphing. Shown below are five images from a morphed sequence.

Figure 4: An original image of the actor Paul Newman (left), and a digitally re-touched image of a younger Newman (right).

2.3 Re-touched

The term morphing has been applied by non-specialist to refer to a broader class of digital tampering which I will refer to as re-touching. Shown in Figure 4, for example, is an original image of the actor Paul Newman, and a digitally re-touched younger Newman. This tampering involved lowering the hairline, removing wrinkles, and removing the darkness under the eyes. These manipulations were a simple matter of copy and pasting small regions from within the same image – e.g., the wrinkles were removed by duplicating wrinkle-free patches of skin onto the wrinkled regions. While this form of tampering can, in the hands of an experienced user, shave a few years off of a person's appearance, it cannot create the radical changes in facial structure needed to produce a, for example, 12-year old Newman.

Figure 5: An original image (top left) and the image enhanced to alter the color (top right), contrast (bottom left) and blur of the background cars (bottom right). The original image was downloaded from `freefoto.com`.

2.4 Enhanced

Shown in Figure 5 are an original image (top left), and three examples of image enhancement: (1) the blue motorcycle was changed to cyan and the red van in the background was changed to yellow; (2) the contrast of the entire image was increased, making the image appear to have been photographed on a bright sunny day; (3) the parked cars were blurred creating a narrower depth of focus as might occur when photographing with a wide aperture. This type of manipulation, unlike compositing, morphing or re-touching is often no more than a few mouse clicks away in Photoshop.

While this type of tampering cannot fundamentally alter the appearance or meaning of an image (as with compositing, morphing and re-touching), it can still have a subtle effect on the interpretation of an image – for example, simple enhancements can obscure or exaggerate image details, or alter the time of day in which the image appears to have been taken.

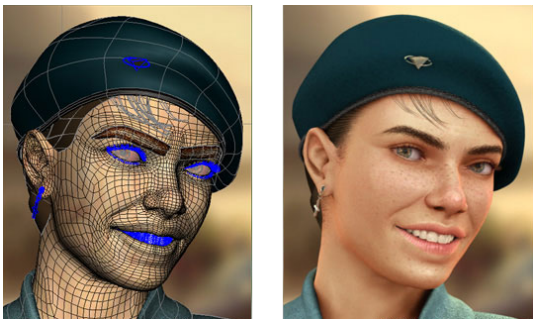

Figure 6: A computer generated model (left) and the resulting rendered image (right), by Alceu Baptistão.

2.5 Computer Generated

Composited, morphed, re-touched and enhanced images, as described in the previous sections, share the property that they typically alter the appearance of an actual photograph (either from a digital camera, or a film camera that was then digitally scanned). Computer generated images, in contrast, are generated entirely by a computer and a skilled artist/programmer (see Sections 6.1 and 6.2 for possible exceptions to this). Such images are generated by first constructing a three-dimensional model of an object (or person) that embodies the desired object shape. The model is then augmented to include color and texture. This complete model is then illuminated with a virtual light source(s), which can approximate a range of indoor or outdoor lighting conditions. This virtual scene is then rendered through a virtual camera to create a final image. Shown in Figure 6, for example, is a partially textured three-dimensional model of a person's head and the final rendered image.

Once the textured three-dimensional model is constructed, an image of the same object from any viewpoint can be easily generated (e.g., we could view the character in Figure 6 from the side, above, or behind). Altering the pose of the object, however, is more involved as it requires changing the underlying three-dimensional model (e.g., animating the character in Figure 6 to nod her head requires updating the model for each head pose – realistic animation of the human form, however, is very difficult, see Section 6.3).

2.6 Painted

Starting with a blank screen, photo-editing software, such as Photoshop, allows a user to create digital works of art, similar to the way a painter would paint or draw on a traditional canvas. Unlike the forms of tampering described in the previous sections, this technique requires a high degree of artistic and technical talent, is very time consuming and is unlikely to yield particularly realistic images.

3 The Child Pornography Prevention Act

The 1996 Child Pornography Prevention Act (CPPA) extended the existing federal criminal laws against child pornography to include certain types of digital images [1]. The CPPA banned, in part, two different forms of "virtual porn":

> §2256(8) child pornography means any visual depiction, including any photograph, film, video, picture, or computer or computer-generated image or picture, whether made or produced by electronic, mechanical, or other means, of sexually explicit conduct, where:
>
> > (B) such visual depiction is, or appears to be, of a minor engaging in sexually explicit conduct;
> >
> > (C) such visual depiction has been created, adapted, or modified to appear that an identifiable minor is engaging in sexually explicit conduct;

Part (B) bans virtual child pornography that appears to depict minors but was produced by means other than using real children, e.g., computer generated images, Section 2.5. Part (C) prohibits virtual child pornography generated using the more common compositing and morphing techniques, Sections 2.1 and 2.2.

Composited, morphed and re-touched images typically originate in photographs of an actual individual. Computer generated images, on the other hand, typically are generated in their entirety from a computer model and thus do not depict any actual individual (see Sections 6.1 and 6.2 for possible exceptions to this). As we will see next, this distinction was critical to the United States Supreme Court's consideration of the constitutionality of the CPPA.

4 Ashcroft v. Free Speech Coalition

In 2002, the United States Supreme Court considered the constitutionality of the CPPA in Ashcroft v. Free Speech Coalition [2]. In their 6-3 ruling, the Court found

that portions of the CPPA, being overly broad and restrictive, violated First Amendment rights. Of particular importance was the Courts ruling on the different forms of "virtual porn" as described above. With respect to §2256(8)(B) the Court wrote, in part:

> Virtual child pornography is not "intrinsically related" to the sexual abuse of children. While the Government asserts that the images can lead to actual instances of child abuse, the causal link is contingent and indirect. The harm does not necessarily follow from the speech, but depends upon some unquantified potential for subsequent criminal acts.

and went on to strike down this provision. With respect to §2256(8)(C) the Court wrote, in part:

> Although morphed images may fall within the definition of virtual child pornography, they implicate the interests of real children and are in that sense closer to the images in Ferber [1] Respondents do not challenge this provision, and we do not consider it.

thus allowing this provision to stand. The Court, therefore, ruled that images containing a minor or portions of a minor are not protected, while computer generated depicting a fictitious minor is constitutionally protected.

With respect to the various forms of digital tampering outlined in Section 2, we need to consider the extent to which painted, computer generated and certain types of morphed and re-touched images can appear, to the casual eye, as authentic. We will also consider what technology is available to expose such images.

5 Is it Real or Virtual?

While the technology to alter digital media is developing at break-neck speeds, the technology to contend with the ramifications is lagging seriously behind. My students and I have been, for the past few years, developing a number of mathematical and computational tools to detect various types of tampering in digital media. Below I review some of this work (see also Appendix A).

[1] In New York v. Ferber (1982), United States Supreme Court upheld a New York statute prohibiting the production, exhibition or selling of any material that depicts any performance by a child under the age of 16 that includes "actual or simulated sexual intercourse, deviate sexual intercourse, sexual bestiality, masturbation, sado-masochistic abuse or lewd exhibitions of the genitals."

5.1 Morphed & Re-touched

We have developed some general techniques that can be used to detect certain types of morphed and re-touched [2] images. These tools work best on high-quality and high-resolution digital images. I only briefly describe these techniques below – see the referenced full papers for complete details.

5.1.1 Color Filter Array

Most digital cameras capture color images using a single sensor in conjunction with an array of color filters. As a result, only one third of the samples in a color image are captured by the camera, the other two thirds being interpolated. This interpolation introduces specific correlations between the samples of a color image. When morphing or re-touching an image these correlations may be destroyed or altered. We have described the form of these correlations, and developed a method that quantifies and detects them in any portion of an image [11]. We have shown the general effectiveness of this technique in detecting traces of digital tampering, and analyzed its sensitivity and robustness to simple counter-attacks.

5.1.2 Duplication

A common manipulation when altering an image is to copy and paste portions of the image to conceal a person or object in the scene. If the splicing is imperceptible, little concern is typically given to the fact that identical (or virtually identical) regions are present in the image. We have developed a technique that can efficiently detect and localize duplicated regions in an image [8]. This technique works by first applying a principal component analysis (PCA) on small fixed-size image blocks to yield a reduced dimension representation. This representation is robust to minor variations in the image due to additive noise or lossy compression. Duplicated regions are then detected by lexicographically sorting all of the image blocks. We have shown the efficacy of this technique on credible forgeries, and quantified its robustness and sensitivity to additive noise and lossy JPEG compression.

[2] In the hands of an experienced user digital re-touching, as shown in Figure 4, can shave a few years off of a person's appearance. These same techniques cannot, however transform an adult (e.g., $25-50$ years old) into a child (e.g., $4-12$ years old). These techniques are simply insufficient to create the radical changes in facial and body structure needed to produce such a young child. Even though such images are currently constitutionally protected, it is unreasonable, given the current technology, to assume that such images do not depict actual minors.

5.2 Computer Generated

Computer graphics rendering software is capable of generating highly photorealistic images that are often very difficult to differentiate from photographic images. We have, however, developed a method for differentiating between photographic and computer generated (photorealistic) images. Specifically, we have shown that a statistical model based on first- and higher-order wavelet statistics reveals subtle but significant differences between photographic and photorealistic images [3]. I will review this technique below, and direct the interested reader to [4, 7] for more details.

5.2.1 Statistical Model

We begin with an image decomposition based on separable quadrature mirror filters (QMFs). The decomposition splits the frequency space into multiple orientations (a vertical, a horizontal and a diagonal subband) and scales. For a color (RGB) image, the decomposition is applied independently to each color channel. The resulting vertical, horizontal and diagonal subbands for scale i are denoted as $V_i^c(x,y)$, $H_i^c(x,y)$, and $D_i^c(x,y)$ respectively, where $c \in \{r,g,b\}$.

The first component of the statistical model consists of the first four order statistics (mean, variance, skewness and kurtosis) of the subband coefficient histograms at each orientation, scale and color channel. While these statistics describe the basic coefficient distributions, they are unlikely to capture the strong correlations that exist across space, orientation and scale. For example, salient image features such as edges tend to orient spatially in certain direction and extend across multiple scales. These image features result in substantial local energy across many scales, orientations and spatial locations. The local energy can be roughly measured by the magnitude of the QMF decomposition coefficients. As such, a strong coefficient in a horizontal subband may indicate that its left and right spatial neighbors in the same subband will also have a large value. Similarly, if there is a coefficient with large magnitude at scale i, it is also very likely that its "parent" at scale $i+1$ will also have a large magnitude.

In order to capture some of these higher-order statistical correlations, we collect a second set of statistics that are based on the errors in a linear predictor of coefficient magnitude. For the purpose of illustration, consider first a vertical band of the green channel at scale i, $V_i^g(x,y)$. A linear predictor for the magnitude of these coefficients in a subset of all possible spatial, orientation, scale and color neighbors is given by:

$$\begin{aligned} |V_i^g(x,y)| &= w_1|V_i^g(x-1,y)| + w_2|V_i^g(x+1,y)| \\ &+ w_3|V_i^g(x,y-1)| + w_4|V_i^g(x,y+1)| \\ &+ w_5|V_{i+1}^g(x/2,y/2)| + w_6|D_i^g(x,y)| \\ &+ w_7|D_{i+1}^g(x/2,y/2)| + w_8|V_i^r(x,y)| \\ &+ w_9|V_i^b(x,y)|, \end{aligned} \tag{1}$$

where $|\cdot|$ denotes absolute value and w_k are the weights. This linear relationship can be expressed more compactly in matrix form as:

$$\vec{v} = Q\vec{w}, \tag{2}$$

where $\vec{v}$ contains the coefficient magnitudes of $V_i^g(x,y)$ strung out into a column vector, the columns of the matrix Q contain the neighboring coefficient magnitudes as specified in Equation (1), and $\vec{w} = (w_1 \ \ldots \ w_9)^T$. The weights $\vec{w}$ are determined by minimizing the following quadratic error function:

$$E(\vec{w}) = [\vec{v} - Q\vec{w}]^2. \tag{3}$$

This error function is minimized by differentiating with respect to $\vec{w}$:

$$\frac{dE(\vec{w})}{d\vec{w}} = 2Q^T(\vec{v} - Q\vec{w}), \tag{4}$$

setting the result equal to zero, and solving for $\vec{w}$ to yield:

$$\vec{w} = (Q^TQ)^{-1}Q^T\vec{v}, \tag{5}$$

Given the large number of constraints (one per pixel) in only nine unknowns, it is generally safe to assume that the 9×9 matrix Q^TQ will be invertible.

Given the linear predictor, the log error between the actual coefficient and the predicted coefficient magnitudes is:

$$\vec{p} = \log(\vec{v}) - \log(|Q\vec{w}|), \tag{6}$$

where the $\log(\cdot)$ is computed point-wise on each vector component. This log error quantifies the correlation of a subband with its neighbors. The mean, variance, skewness and kurtosis of this error are collected to characterize its distribution. This process is repeated for scales $i = 1, \ldots, n-1$, and for the subbands V_i^r and V_i^b, where the linear predictors for these subbands are of the form:

$$\begin{aligned} |V_i^r(x,y)| &= w_1|V_i^r(x-1,y)| + w_2|V_i^r(x+1,y)| \\ &+ w_3|V_i^r(x,y-1)| + w_4|V_i^r(x,y+1)| \\ &+ w_5|V_{i+1}^r(x/2,y/2)| + w_6|D_i^r(x,y)| \\ &+ w_7|D_{i+1}^r(x/2,y/2)| + w_8|V_i^g(x,y)| \\ &+ w_9|V_i^b(x,y)|, \end{aligned} \tag{7}$$

[3] We have used a similar technique to detect messages hidden within digital images (steganography) [3, 5, 6].

and

$$\begin{aligned}|V_i^b(x,y)| &= w_1|V_i^b(x-1,y)| + w_2|V_i^b(x+1,y)| \\ &+ w_3|V_i^b(x,y-1)| + w_4|V_i^b(x,y+1)| \\ &+ w_5|V_{i+1}^b(x/2,y/2)| + w_6|D_i^b(x,y)| \\ &+ w_7|D_{i+1}^b(x/2,y/2)| + w_8|V_i^r(x,y)| \\ &+ w_9|V_i^g(x,y)|. \end{aligned} \quad (8)$$

A similar process is repeated for the horizontal and diagonal subbands. As an example, the predictor for the green channel takes the form:

$$\begin{aligned}|H_i^g(x,y)| &= w_1|H_i^g(x-1,y)| + w_2|H_i^g(x+1,y)| \\ &+ w_3|H_i^g(x,y-1)| + w_4|H_i^g(x,y+1)| \\ &+ w_5|H_{i+1}^g(x/2,y/2)| + w_6|D_i^g(x,y)| \\ &+ w_7|D_{i+1}^g(x/2,y/2)| + w_8|H_i^r(x,y)| \\ &+ w_9|H_i^b(x,y)|, \end{aligned} \quad (9)$$

and

$$\begin{aligned}|D_i^g(x,y)| &= w_1|D_i^g(x-1,y)| + w_2|D_i^g(x+1,y)| \\ &+ w_3|D_i^g(x,y-1)| + w_4|D_i^g(x,y+1)| \\ &+ w_5|D_{i+1}^g(x/2,y/2)| + w_6|H_i^g(x,y)| \\ &+ w_7|V_i^g(x,y)| + w_8|D_i^r(x,y)| \\ &+ w_9|D_i^b(x,y)|. \end{aligned} \quad (10)$$

For the horizontal and diagonal subbands, the predictor for the red and blue channels are determined in a similar way as was done for the vertical subbands, Equations (7)-(8). For each oriented, scale and color subband, a similar error metric, Equation(6), and error statistics are computed.

For a multi-scale decomposition with scales $i = 1, ..., n$, the total number of basic coefficient statistics is $36(n-1)$ ($12(n-1)$ per color channel), and the total number of error statistics is also $36(n-1)$, yielding a grand total of $72(n-1)$ statistics. These statistics form the feature vector to be used to discriminate between photographic and photorealistic images.

5.2.2 Classification

From the measured statistics of a training set of images labeled as photographic or photorealistic, our goal is to build a classifier that can determine to which category a novel test image belongs. To this end, a support vector machine (SVM) is employed. I will briefly describe, in increasing complexity, three classes of SVMs. The first, linear separable case is mathematically the most straight-forward. The second, linear non-separable case, contends with situations in which a solution cannot be found in the former case. The third, non-linear case, affords the most flexible classification scheme and often gives the best classification accuracy.

Linear Separable SVM: Denote the tuple $(\vec{x}_i, y_i)$, $i = 1, ..., N$ as exemplars from a training set of photographic and photorealistic images. The column vector $\vec{x}_i$ contains the measured image statistics as outlined in the previous section, and $y_i = -1$ for photorealistic images, and $y_i = 1$ for photographic images. The linear separable SVM classifier amounts to a hyperplane that separates the positive and negative exemplars. Points which lie on the hyperplane satisfy the constraint:

$$\vec{w}^t\vec{x}_i + b = 0, \quad (11)$$

where $\vec{w}$ is normal to the hyperplane, $|b|/||\vec{w}||$ is the perpendicular distance from the origin to the hyperplane, and $||\cdot||$ denotes the Euclidean norm. Define now the margin for any given hyperplane to be the sum of the distances from the hyperplane to the nearest positive and negative exemplar. The separating hyperplane is chosen so as to maximize the margin. If a hyperplane exists that separates all the data then, within a scale factor:

$$\vec{w}^t\vec{x}_i + b \geq 1, \qquad \text{if } y_i = 1 \quad (12)$$

$$\vec{w}^t\vec{x}_i + b \leq -1, \qquad \text{if } y_i = -1. \quad (13)$$

These pair of constraints can be combined into a single set of inequalities:

$$(\vec{w}^t\vec{x}_i + b)\, y_i - 1 \geq 0, \qquad i = 1, ..., N. \quad (14)$$

For any given hyperplane that satisfies this constraint, the margin is $2/||\vec{w}||$. We seek, therefore, to minimize $||\vec{w}||^2$ subject to the constraints in Equation (14).

For largely computational reasons, this optimization problem is reformulated using Lagrange multipliers, yielding the following Lagrangian:

$$\begin{aligned}L(\vec{w}, b, \alpha_1, ..., \alpha_N) &= \frac{1}{2}||\vec{w}||^2 \\ &- \sum_{i=1}^{N} \alpha_i \,(\vec{w}^t\vec{x}_i + b)\, y_i \\ &+ \sum_{i=1}^{N} \alpha_i, \end{aligned} \quad (15)$$

where α_i are the positive Lagrange multipliers. This error function should be minimized with respect to $\vec{w}$ and b, while requiring that the derivatives of $L(\cdot)$ with respect to each α_i is zero and constraining $\alpha_i \geq 0$, for all i. Because this is a convex quadratic programming problem, a solution to the dual problem yields

the same solution for $\vec{w}$, b, and $\alpha_1, ..., \alpha_N$. In the dual problem, the same error function $L(\cdot)$ is maximized with respect to α_i, while requiring that the derivatives of $L(\cdot)$ with respect to $\vec{w}$ and b are zero and the constraint that $\alpha_i \geq 0$. Differentiating with respect to $\vec{w}$ and b, and setting the results equal to zero yields:

$$\vec{w} = \sum_{i=1}^{N} \alpha_i \vec{x}_i y_i \quad (16)$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0. \quad (17)$$

Substituting these equalities back into Equation (15) yields:

$$L_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \vec{x}_i^t \vec{x}_j y_i y_j. \quad (18)$$

Maximization of this error function may be realized using any of a number of general purpose optimization packages that solve linearly constrained convex quadratic problems.

A solution to the linear separable classifier, if it exists, yields values of α_i, from which the normal to the hyperplane can be calculated as in Equation (16), and from the Karush-Kuhn-Tucker (KKT) condition:

$$b = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \vec{w}^t \vec{x}_i \right), \quad (19)$$

for all i, such that $\alpha_i \neq 0$. From the separating hyperplane, $\vec{w}$ and b, a novel exemplar, $\vec{z}$, can be classified by simply determining on which side of the hyperplane it lies. If the quantity $\vec{w}^t \vec{z} + b$ is greater than or equal to zero, then the exemplar is classified as photographic, otherwise the exemplar is classified as photorealistic.

Linear Non-Separable SVM: It is possible, and even likely, that the linear separable SVM will not yield a solution when, for example, the training data do not uniformly lie on either side of a separating hyperplane. Such a situation can be handled by softening the initial constraints of Equation (12) and (13). Specifically, these constraints are modified with "slack" variables, ξ_i, as follows:

$$\vec{w}^t \vec{x}_i + b \geq 1 - \xi_i, \quad \text{if } y_i = 1 \quad (20)$$

$$\vec{w}^t \vec{x}_i + b \leq -1 + \xi_i, \quad \text{if } y_i = -1, \quad (21)$$

with $\xi_i \geq 0,\ i = 1, ..., N$. A training exemplar which lies on the "wrong" side of the separating hyperplane will have a value of ξ_i greater than unity. We seek a hyperplane that minimizes the total training error, $\sum_i \xi_i$, while still maximizing the margin. A simple error function to be minimized is $||\vec{w}||^2/2 + C \sum_i \xi_i$, where C is a user selected scalar value, whose chosen value controls the relative penalty for training errors. Minimization of this error is still a quadratic programming problem. Following the same procedure as the previous section, the dual problem is expressed as maximizing the error function:

$$L_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \vec{x}_i^t \vec{x}_j y_i y_j, \quad (22)$$

with the constraint that $0 \leq \alpha_i \leq C$. Note that this is the same error function as before, Equation (18) with the slightly different constraint that α_i is bounded above by C. Maximization of this error function and computation of the hyperplane parameters are accomplished as described in the previous section.

Non-Linear SVM: Fundamental to the SVMs outlined in the previous two sections is the limitation that the classifier is constrained to a linear hyperplane. It is often the case that a non-linear separating surface greatly improves classification accuracy. Non-linear SVMs afford such a classifier by first mapping the training exemplars into a higher (possibly infinite) dimensional Euclidean space in which a linear SVM is then employed. Denote this mapping as:

$$\Phi : \mathcal{L} \rightarrow \mathcal{H}, \quad (23)$$

which maps the original training data from $\mathcal{L}$ into $\mathcal{H}$. Replacing $\vec{x_i}$ with $\Phi(\vec{x_i})$ everywhere in the training portion of the linear separable or non-separable SVMs of the previous sections yields an SVM in the higher-dimensional space $\mathcal{H}$.

It can, unfortunately, be quite inconvenient to work in the space $\mathcal{H}$ as this space can be considerably larger than the original $\mathcal{L}$, or even infinite. Note, however, that the error function of Equation (22) to be maximized depends only on the inner products of the training exemplars, $\vec{x}_i^t \vec{x}_j$. Given a "kernel" function such that:

$$K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i)^t \Phi(\vec{x}_j), \quad (24)$$

an explicit computation of Φ can be completely avoided. There are several choices for the form of the kernel function, for example, radial basis functions or polynomials. Replacing the inner products $\Phi(\vec{x}_i)^t \Phi(\vec{x}_j)$ with the kernel function $K(\vec{x}_i, \vec{x}_j)$ yields an SVM in the space $\mathcal{H}$ with minimal computational impact over working in the original space $\mathcal{L}$.

With the training stage complete, recall that a novel exemplar, $\vec{z}$, is classified by determining on which side of the separating hyperplane (specified by $\vec{w}$ and b) it lies. Specifically, if the quantity $\vec{w}^t\Phi(\vec{z}) + b$ is greater than or equal to zero, then the exemplar is classified as photographic, otherwise the exemplar is classified photorealistic. The normal to the hyperplane, $\vec{w}$, of course now lives in the space $\mathcal{H}$, making this testing impractical. As in the training stage, the classification can again be performed via inner products. From Equation (16):

$$\begin{aligned} \vec{w}^t\Phi(\vec{z}) + b &= \sum_{i=1}^{N} \alpha_i \Phi(\vec{x}_i)^t \Phi(\vec{z}) y_i + b \\ &= \sum_{i=1}^{N} \alpha_i K(\vec{x}_i, \vec{z}) y_i + b. \end{aligned} \quad (25)$$

Thus both the training and classification can be performed in the higher-dimensional space, affording a more flexible separating hyperplane and hence better classification accuracy. We next show the performance of a non-linear SVM in the classification of images as photographic or photorealistic. The SVMs classify images based on the statistical feature vector as described in Section 5.2.1

5.2.3 Results

We have constructed a database of $40,000$ photographic and $6,000$ photorealistic images[4]. All of the images consist of a broad range of indoor and outdoor scenes, and the photorealistic images were rendered using a number of different software packages (e.g., 3D Studio Max, Maya, SoftImage 3D, PovRay, Lightwave 3D and Imagine). All of the images are color (RGB), JPEG compressed (with an average quality of 90%), and typically on the order of 600×400 pixels in size.

From this database of 46,000 images, statistics as described in Section 5.2.1 were extracted. To accommodate different image sizes, only the central 256×256 region of each image was considered. For each image region, a four-level three-orientation QMF pyramid was constructed for each color channel, from which a 216-dimensional feature vector (72 per color channel) of coefficient and error statistics was collected.

From the 46,000 feature vectors, 32,000 photographic and 4,800 photorealistic feature vectors were used to train a non-linear SVM. The remaining feature vectors were used to test the classifier. In the results presented here, the training/testing split was done randomly – the average testing classification accuracy over 100 such splits is reported. With a 0.5% false-negative rate (a photorealistic image mis-classified as photographic), the SVM correctly correctly classified 72% of the photographic images.

[4]The photographic images were downloaded from `www.freefoto.com`, the photorealistic images were downloaded from `www.raph.com` and `www.irtc.org`.

5.3 Painted

As described in Section 2.6, realistic digitally painted images are extremely difficult to create. I believe, nevertheless, that the technique described in the previous section for differentiating between photographic and computer generated images will also able to differentiate between photographic and painted images. I have not, however, tested this directly for lack of the appropriate data (i.e., realisticly painted images).

6 Modern Technology

As computer and imaging technology continues to develop, the distinction between real and virtual will become increasingly more difficult to make. These days computer generated images, as described in Section 2.5, are typically generated entirely within the confines of a computer and the imagination of the artist/programmer. The creation of these images, therefore, do not involve photographs of actual people. With respect to child pornography, such images are currently constitutionally protected.

I describe below three emerging technologies that employ people in various stages of computer generated imaging. These technologies have emerged in order to allow for the creation of more realistic images. While not yet readily available, I believe that these technologies will eventually make it increasingly more difficult to determine if a person was involved in any stage of the creation of a computer generated image.

6.1 Image-Based Rendering

The appearance of computer generated images, as described in Section 2.5, is typically dictated by the color and texture that is mapped onto the virtual object or person. Image-based rendering allows for actual photographs to be mapped directly onto the rendered object or person. For example, in creating a virtual person, the three-dimensional model may be created by the computer, and a photograph of a real person overlayed onto that model. The resulting rendered image is part real and part virtual – though the underlying three-dimensional shape of the person is virtual, the

Figure 8: The 3-D motion capture system by Meta-Motion.

image may contain recognizable features of a real person.

6.2 3-D Laser Scanning

Computer generated images, as described in Section 2.5, are generated by first constructing a three-dimensional model of an object or person. Computer graphics software provides a number of convenient tools to aid in creating such models.

A number of commercially available scanners directly generate a three-dimensional model of an actual object or person. The Cyberware whole body scanner, for example, captures (in less than 17 seconds) the three-dimensional shape and color of the entire human body, Figure 7. Shown in this figure is the scanner and five views of a full-body scan. This scanned model can then be imported into a computer graphics software and texture mapped as desired. The resulting rendered image is part real and part virtual – though the image may not be recognized as a real person, the underlying three-dimensional model is that of a real person.

6.3 3-D Motion Capture

If you have seen any computer animated movie (e.g., Toy Story, Shrek, etc.), you may have noticed that the motion of human characters often looks stilted and awkward. At least two reasons for this are that the biomechanics of human motion are very difficult to model and recreate, and we seem to have an intensely acute sensitivity to human motion (e.g., long before we can clearly see their face, we can often recognize a familiar person from their gait).

A number of commercially available systems can capture the three-dimensional motion of a human as they undergo complex motions. Motion capture systems, such as that shown in Figure 8, measure the three-dimensional position of several key points on the human body, typically at the joints. This data can then be used to animate a completely computer generated person. The resulting animation is part real and part virtual – while the character does not depict a real person, the underlying motion is that of a real person.

7 Discussion

Today's technology allows digital media to be altered and manipulated in ways that were simply impossible 20 years ago. Tomorrow's technology will almost certainly allow for us to manipulate digital media in ways that today seem unimaginable. And as this technology continues to evolve it will become increasingly more difficult for the courts, the media and, in general, the average person to keep pace with understanding its power and it limits.

I have tried in this report to review some of the current digital technology that allows for images to be created and altered, with the hope that it will help the courts and others grapple with some difficult technical and legal issues currently facing us. It is also my hope that the mathematical and computational techniques that we have developed (and continue to develop) will help the courts contend with this exciting and at times puzzling digital age.

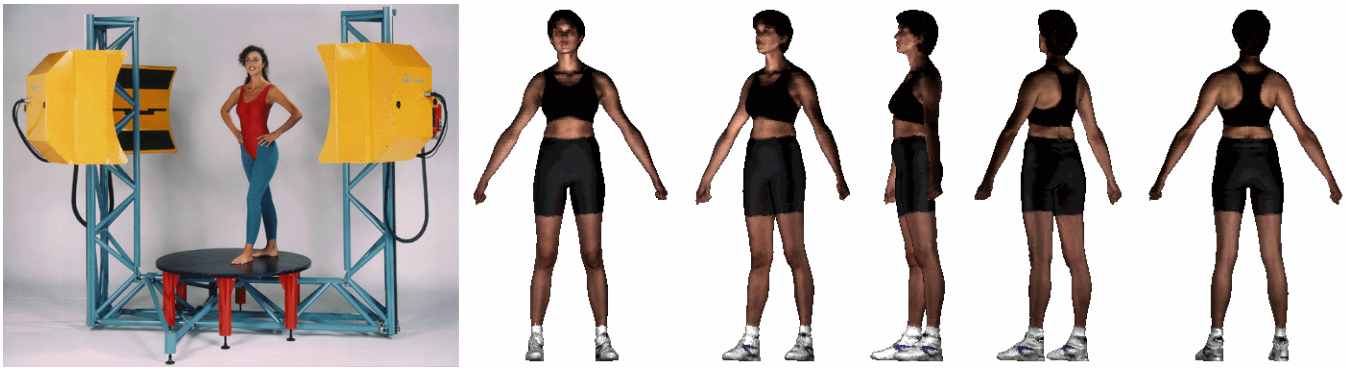

Figure 7: The whole body 3-D scanner by Cyberware and five views of a full-body scan.

A Exposing Digital Composites

In addition to the development of a technique to differentiate between photographic and computer generated images, we have developed techniques to detect traces of tampering in photographic images that would result from digital image compositing, Section 2.1. These approaches work on the assumption that although digital forgeries may leave no visual clues of having been tampered with, they may, nevertheless, alter the underlying statistics of an image. Described below are four techniques for detecting various forms of digital tampering (also applicable are the two techniques described in Section 5.1). Provided are only brief descriptions – see the referenced full papers for complete details.

A.1 Re-Sampling

Consider the creation of a digital image that shows a pair of famous movie stars, rumored to have a romantic relationship, walking hand-in-hand. Such an image could be created by splicing together individual images of each movie star and overlaying the digitally created composite onto a sunset beach. In order to create a convincing match, it is often necessary to re-size, rotate, or stretch portions of the images. This process requires re-sampling the original image onto a new sampling lattice. Although this re-sampling is often imperceptible, it introduces specific correlations into the image, which when detected can be used as evidence of digital tampering. We have described the form of these correlations and how they can be automatically detected in any portion of an image [9]. We have shown the general effectiveness of this technique and analyzed its sensitivity and robustness to simple counter-attacks.

A.2 Double JPEG Compression

Tampering with a digital image requires the use of a photo-editing software such as Adobe PhotoShop. In the making of digital forgeries an image is loaded into the editing software, some manipulations are performed and the image is re-saved. Since most images are stored in JPEG format (e.g., a majority of digital cameras store images directly in JPEG format), it is likely that both the original and forged images are stored in this format. Notice that in this scenario the forged image is double JPEG compressed. Double JPEG compression introduces specific artifacts not present in singly compressed images [10]. These artifacts can be used as evidence of digital tampering. Note, however, that double JPEG compression does not necessarily prove malicious tampering. For example, it is possible for a user to simply re-save a high quality JPEG image with a lower quality. The authenticity of a double JPEG compressed image should, nevertheless, be called into question.

A.3 Signal to Noise

Digital images have an inherent amount of noise introduced either by the imaging process or digital compression. The amount of noise is typically uniform across the entire image. If two images with different noise levels are spliced together, or if small amounts of noise are locally added to conceal traces of tampering, then variations in the signal to noise ratio (SNR) across the image can be used as evidence of tampering. Measuring the SNR is non-trivial in the absence of the original signal. We have shown how a *blind* SNR estimators can be employed to locally measure noise variance [10]. Differences in the noise variance across the image can be used as evidence of digital tampering.

A.4 Gamma Correction

In order to enhance the perceptual quality of digital images, digital cameras often introduce some form of luminance non-linearity. The parameters of this non-linearity are usually dynamically chosen and depend on the camera and scene dynamics — these parameters are, however, typically held constant within an image. The presence of several distinct non-linearities in an image is a sign of possible tampering. For example, imagine a scenario where two images are spliced together. If the images were taken with different cameras or in different lightning conditions, then it is likely that different non-linearities are present in the composite image. It is also possible that local non-linearities are applied in the composite image in order to create a convincing luminance match. We have shown that a non-linear transformation introduces specific correlations in the Fourier domain [10]. These correlations can be detected and estimated using tools from polyspectral analysis. This technique is employed to detect if an image contains multiple non-linearities, as might result from digital tampering.

References

[1] http://www.cs.dartmouth.edu/farid/publications/cppa96.html.

[2] http://www.cs.dartmouth.edu/farid/publications/ashcroft.v.freespeechcoalition.pdf.

[3] H. Farid. Detecting hidden messages using higher-order statistical models. In *International Conference on Image Processing*, Rochester, New York, 2002.

[4] H. Farid and S. Lyu. Higher-order wavelet statistics and their application to digital forensics. In *IEEE Workshop on Statistical Analysis in Computer Vision*, Madison, Wisconsin, 2003.

[5] S. Lyu and H. Farid. Detecting hidden messages using higher-order statistics and support vector machines. In *5th International Workshop on Information Hiding*, 2002.

[6] S. Lyu and H. Farid. Steganalysis using color wavelet statistics and one-class support vector machines. In *SPIE Symposium on Electronic Imaging*, 2004.

[7] S. Lyu and H. Farid. How realistic is photorealistic? *IEEE Transactions on Signal Processing*, 2005. (in press).

[8] A.C. Popescu and H. Farid. Exposing digital forgeries by detecting duplicated image regions. Technical Report TR2004-515, Department of Computer Science, Dartmouth College, 2004.

[9] A.C. Popescu and H. Farid. Exposing digital forgeries by detecting traces of re-sampling. *IEEE Transactions on Signal Processing*, 2004. (in press).

[10] A.C. Popescu and H. Farid. Statistical tools for digital forensics. In *Proceedings of the 6th Information Hiding Workshop*, May 2004.

[11] A.C. Popescu and H. Farid. Exposing digital forgeries in color filter array interpolated images. *IEEE Transactions on Signal Processing*, 2005. (in review).