A Reliable and Reversible Image Privacy Protection Based on False Colors

Serdar Çiftçi, Ahmet Oğuz Akyüz, and Touradj Ebrahimi, *Member*, *IEEE*

***Abstract*—Protection of visual privacy has become an indispensable component of video surveillance systems due to pervasive use of video cameras for surveillance purposes. In this paper, we propose two fully reversible privacy protection schemes implemented within the JPEG architecture. In both schemes, privacy protection is accomplished by using false colors with the first scheme being adaptable to other privacy protection filters while the second is false color specific. Both schemes support either a lossless mode in which the original unprotected content can be fully extracted or a lossy mode, which limits file size while still maintaining intelligibility. Our method is not region-of-interest (ROI) based and can be applied on entire frames without compromising intelligibility. This frees the user from having to define ROIs and improves security as tracking ROIs under dynamic content may fail, exposing sensitive information. Our experimental results indicate the favorability of our method over other commonly used solutions to protect visual privacy.**

***Index Terms*—Privacy protection, false color, JPEG.**

I. Introduction

SECURITY NEEDS have become an indispensable part of everyday life and video surveillance is one of the most resorted solutions to address such concerns. A recent report by the British Security Industry Authority reveals that there are almost 6 million CCTV cameras in Britain alone, or around one camera for every 11 individuals [1]. This excessive usage of visual surveillance raises public concerns about individuals' privacy. In addition to surveillance, video cameras are now being routinely used in ambient-assisted living applications [2], in which ensuring visual privacy is also a critical concern [3].

Due to difficulties of storing and analyzing this huge amount of multimedia data in local servers, deferring these tasks to the cloud servers has gained popularity [4]. However, this further exacerbates privacy concerns as such data can also be acquired by unauthorized parties. This gave rise to various data hiding schemes in which the data stored in the servers is encrypted in a reversible manner, either as a ciphertext [5] or as plaintext [6].

However, none of the existing privacy-protection methods for multimedia content seem to have gained popularity. One of the reasons behind this is that all visual privacy protection solutions proposed so-far rely on either manual identification of sensitive regions or require a computer vision module to do so, resulting in complex operations that often lack robustness and therefore reliability [7]. JPEG-based cloud security solutions are also improper for monitoring tasks as they either completely scramble or map the content to a different target image, which is unrelated to the original [6].

Here, we propose a new approach to visual privacy protection that offers all features present in state-of-the-art solutions, while not relying on either manual or automatic sensitive regions detection, hence offering a simple and robust solution to the protection of visual privacy surveillance, monitoring, and multimedia applications.

Our proposed solution aims to strike a balance between various criteria that are important in visual privacy protection, namely *privacy*, *intelligibility*, *reversibility*, *security*, and *robustness*. More specifically our goals are: (1) An individual recorded in a security video should not be easily identifiable by human observers and face recognition algorithms (privacy); (2) the privacy protected video should still allow identification of suspicious behaviors and gathering of non-sensitive information such as the number of people in a given area (intelligibility); (3) in case of a crime, the privacy protected footage could be reversed to obtain the original unprotected footage by authorized users (reversibility); (4) this reversal could only be performed by legally authorized parties and not by any third parties who may have acquired the protected content by some means (security); and (5) privacy protection should be robust in that it should not depend on fragile computer vision algorithms or manual annotations that may fail to detect sensitive regions in some frames (robustness).

To accomplish these goals, we propose two *false color* based schemes implemented within the JPEG architecture. Both schemes are related; however, the first is not specific to false colors and, if desired, can be used with other privacy protection algorithms. The second, on the other hand, is tailored to be used with false colors. The benefit of the second scheme is that it significantly reduces the file size of the protected content by leveraging the coherence between the original image and the false colored version.

Our experimental results involving perceptually meaningful quality metrics and face recognition algorithms indicate the favorability of our method over other privacy protection methods. Furthermore, we validate our technique using a subjective experiment, which confirms that our method achieves a better intelligibility-privacy balance than the compared techniques. Finally, we show that our method is resistant to attacks that aim to recover the original information from the protected content.

S. Çiftçi is with Department of Computer Engineering at METU, Ankara, Turkey. e-mail: sciftci@ceng.metu.edu.tr.

A. O. Akyüz is with Department of Computer Engineering at METU, Ankara, Turkey. e-mail: akyuz@ceng.metu.edu.tr

T. Ebrahimi is with Multimedia Signal Processing Group at EPFL, Lausanne, Switzerland. e-mail: touradj.ebrahimi@epfl.ch

II. Related Work

The goal of visual privacy protection is to prevent sensitive information present in an image (or video) from being revealed to the viewers of this content. Many approaches have been proposed to achieve this goal, varying in complexity from straightforward filtering methods to sophisticated computer vision based algorithms. Two excellent surveys are provided by [7] and [8]. In this section, we first provide an overview of the main categories of visual privacy protection algorithms. We then review false color based privacy protection. Finally, we discuss JPEG and relevant methods that use JPEG-metadata as restorative information.

A. Visual Privacy Protection (VPP)

The three most commonly used VPP approaches are masking, blurring, and pixelation (Figure 1). Masking involves replacing a given ROI with a solid color. Blurring updates each pixel value with a Gaussian average of its neighborhood. Pixelation divides the image into a non-overlapping grid and sets the color of each pixel to the mean color of its enclosing grid cell [9].

There are several problems with these simple techniques. Firstly, they not only distort sensitive information but may also impair the intelligibility of non-sensitive content. Secondly, they are irreversible: even if blurring and pixelation can be reversed to some extent, recovering the original information is generally impossible. Finally, they are vulnerable to certain types of attacks and therefore may fail to fully conceal the identities of recorded individuals [10].

Arguably the most secure type of algorithms are those that involve encryption. These algorithms treat either the entire image or a selected ROI as a bitstream and apply various well-known encryption algorithms such as DES, RC5, AES, or RSA [11]. Due to time complexity of these algorithms, more lightweight encryption methods specific for digital video have been proposed [12].

A related group of algorithms to encryption are those that involve scrambling the video content to make it unrecognizable to viewers [13]. These algorithms permute the data based on a pseudo-random sequence. The original content can be recovered solely with the knowledge of the seed value that gave rise to this sequence. Scrambling can be performed in the spatial domain, transform domain (e.g. frequency domain), or a format-dependent codestream domain. For instance, Dufaux and Ebrahimi [14] propose two scrambling methods for the MPEG-4 format. The first method is based on pseudo-randomly flipping the signs of the AC coefficients during MPEG-4 encoding. The second method takes as input an encoded codestream and tries to identify the relevant syntax elements to perform similar sign-flipping operations. Although these methods are found to be superior to blurring and pixelation for hiding identity [15], [16], they significantly hamper the intelligibility and the pleasantness of protected content.

Another type of algorithms specifically aim to protect the privacy of faces. To this end, they either require the face regions to be manually marked or rely on a face detection algorithm [17] to do so. The most well-known algorithms that belong to this group are the k-Same family of algorithms. They try to anonymize a face by replacing the original face with an average face computed over k number of face images [18]. The utility of these algorithms are improved by later approaches that aim to preserve facial expressions, gender, and overall appearance [19], [20], [21].

Two other notable methods of face anonymization (i.e. de-identification) are known as morphing and warping. In the former, the input face image is morphed to a target face based on an interpolation parameter [22]. The interpolation is performed to steer both the intensity and positions of the key points in the input face toward the target face. In the warping approach, the automatically selected key points are randomly shifted to different positions and the remaining pixels are computed by transformation and interpolation [23]. The drawback of both approaches is that, depending on the interpolation parameter and the warping strength, the original face may be unrecoverable.

If the objects of interest can be accurately identified, abstraction algorithms can be used to replace the actual objects by their abstracted versions. For instance, a human figure can be replaced with a silhouette [24], [25], caricature [26], 3D avatar [27], or a stick-figure [28]. A comparison of several abstraction models is provided by Chimoni et al. [29].

An alternative to abstraction is to completely remove the sensitive objects. The remaining gap is then filled by using image or video inpainting algorithms [30], [31], [32]. It should be noted that these algorithms are computationally very expensive and are generally not suitable for real-time applications [33]. Furthermore, such approaches are not suitable for surveillance or assisted-living applications due to lack of intelligibility.

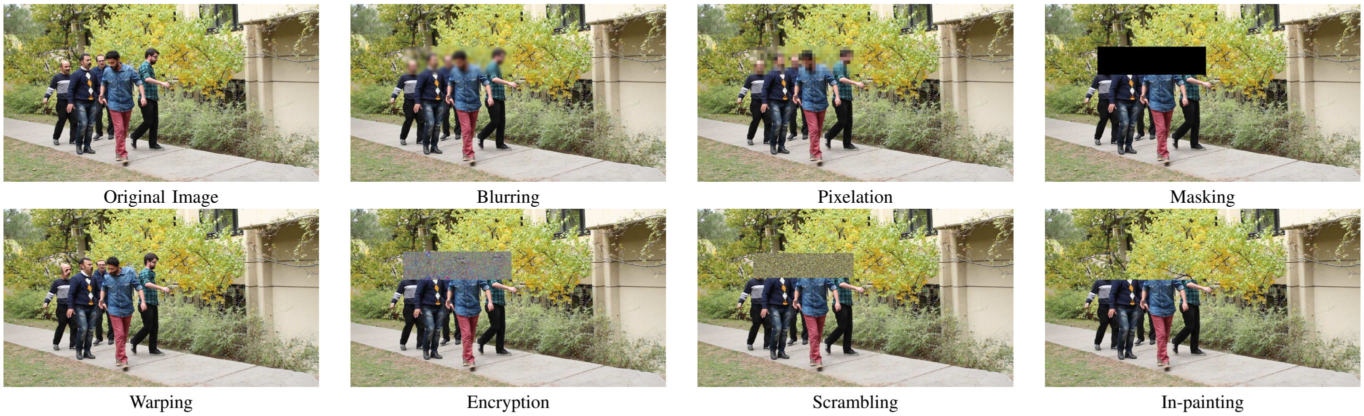
The primary drawback of all of these VPP approaches is that they require either a user-defined or automatically extracted ROI to apply privacy protection. Applying protection on the full frames severely impairs the intelligibility of the captured data. Manually defining a ROI is not practical and the robustness of automatically extracting a ROI depends on the robustness of object/human detection algorithms, which are known to fail especially in harsh capture conditions.

B. False Color Based VPP

In image processing, false colors are typically used as a visualization aid to represent otherwise invisible information. For instance, it is not uncommon to represent high dynamic range (HDR) images in false color to convey the high range of luminances in the captured scene [34].

Recently, false colors are used for the purpose of visual privacy protection. To this end, an RGB input image is first transformed into grayscale. The 8-bit grayscale value is then used to index into an RGB color table (i.e. palette) and the corresponding RGB triplet is used to replace the original pixel value (see Figure 2). This approach has been applied for both images [35] and video [36].

The primary advantage of false color based VPP is that it can be applied on the entire image without compromising intelligibility. In other words, selection of a ROI is not required, which makes this method robust against the fragility


Original Image

Blurring

Pixelation

Masking

Warping

Encryption

Scrambling

In-painting

Fig. 1. Common privacy protection methods.

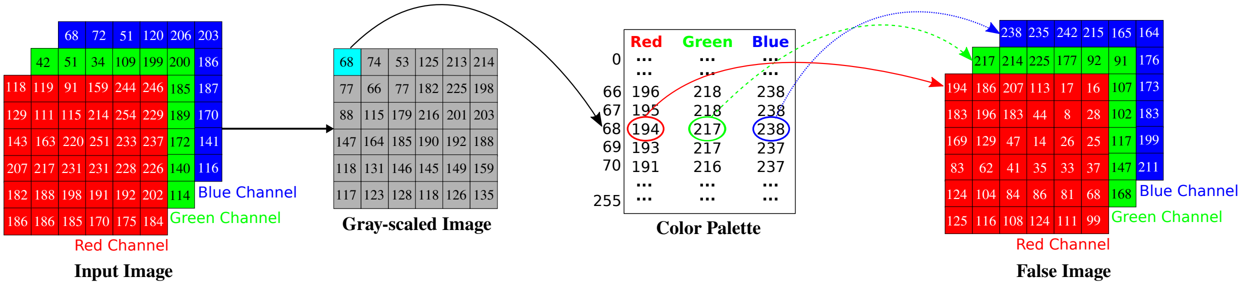


Fig. 2. The basic false coloring technique.

of computer vision algorithms that aim to detect sensitive regions. However, this method is not fully reversible due to two reasons: (1) The original color to grayscale conversion dismisses the color information and (2) the color palettes are typically not one-to-one, which means that two different grayscale values may get mapped to the same color value rendering the recovery of the original grayscale value impossible.

In this paper, we propose to extend the false color based VPP such that the original unprotected content can be perfectly recovered, while still allowing for lossy recovery if the file size is the primary concern. Furthermore, we implement this approach within the industry-standard JPEG format. More specifically, we produce a single JPEG output in which the main image is the protected one with restorative information saved in the metadata. An authorized viewer can then decrypt this extra information to recover the original.

C. Usage of JPEG Metadata Embedding

One of the main strengths of the JPEG standard [37], [38] is its support for metadata embedding mechanism in the form of application markers. These markers can be used to store various forms of metadata such as EXIF and IPTC, or they can be used to store vendor-specific information [39]. Traditionally, this information was used to store extra information that each vendor may want to make available. More recently, however, this metadata has been judiciously used to embed restorative information. This information, when combined with the main image, may be used to expand the color gamut and/or dynamic range of the recorded image or video [40], [41], [42], [43].

The most recent JPEG format, known as JPEGXT, also uses this metadata extensively to store HDR images within a JPEG file according to several profiles [44]. In our method, we also use this metadata to allow recovery of the original image from the protected one.

Of most related to our technique is a recent method called Secure JPEG [45], which scrambles the given ROIs and saves the necessary information to descramble it in an application marker. However, unlike our method, it cannot be used on full frames as it would destroy the intelligibility of the data. It is targeted toward social photo sharing applications, rather than to be used for surveillance or ambient-assisted living tasks.

D. Privacy Models

While most VPP methods implicitly refer to terms such as utility (i.e. intelligibility) and privacy, Saini et al. propose a mathematical definition for them [46], [47]. They define two models, one for privacy loss (Γ) and the other for utility loss (U). Both models are comprised of multiple tasks with each task modeling a different aspect of privacy and utility. The exact definition of tasks are left to the user as different users may consider different tasks to be important for either attribute. The two models are combined to yield the following energy function:

$$E = \eta\Gamma(\mathcal{F}(V)) + (1-\eta)U(\mathcal{F}(V)), \qquad (1)$$

where $\mathcal{F}$ defines a data transformation, V represents the original data, and η is a user parameter used to define the importance of privacy loss over utility loss. Using this model, the authors develop a hybrid global model in which blurring and quantization are used in sequence to achieve a better trade-off between privacy and utility [47].

III. Proposed Method

Our proposed method consists of two schemes. The first scheme is more general in that, besides false coloring, it can be used with any privacy protection strategy. The second scheme, on the other hand, makes use of the color palette to reduce the file size of the protected image without affecting its intelligibility. Both schemes support lossless and lossy modes as discussed in detail in the following subsections.

A. Scheme One

1) Protection Pipeline: The protection pipeline of the first scheme is illustrated in Figure 3. Here, the input image (I) is first converted into grayscale. Next, by using the grayscale values as indices into a color palette, the false color image (FI) is obtained. This image is saved as the main JPEG image in the output file. FI is then JPEG encoded and decoded to simulate what the decoder will see at the decoding end. We call this image FI'. Afterwards, the difference image (DI) is computed as:

$$DI_c(x, y) = |I_c(x, y) - FI'_c(x, y)|, \tag{2}$$

where (x, y) indicates the pixel index and $c \in \{R, G, B\}$. As this difference is sometimes negative, an accompanying sign image (SI) is computed as well:

$$SI_c(x, y) = \begin{cases} 1 & \text{if} \quad I_c(x, y) - FI'_c(x, y) < 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

For efficient storage, we use a single bit for each difference and then compress it losslessly using the zlib compression algorithm [48]. The difference image may be either losslessly or lossily compressed. For lossless compression, we use zlib whereas for lossy compression we use JPEG compression or downsampling (both can be used simultaneously as well). The compressed and encrypted DI and SI are then saved as metadata in JPEG application markers.

2) Recovery Pipeline: For recovery (Figure 4), the JPEG file is first decoded to obtain the false color image, difference image, and the sign image. Note that the decoded false color image will be equal to FI' introduced above. In the lossless mode, the difference and sign images will be equal to DI and SI. These streams are first decrypted using the authorization key and then decompressed. The recovered image, R, is obtained by:

$$R_c(x, y) = FI'_c(x, y) + sDI_c(x, y) \tag{4}$$

with s computed as:

$$s = \begin{cases} 1 & \text{if} \quad SI_c(x, y) = 0, \\ -1 & \text{otherwise.} \end{cases} \tag{5}$$

Note that, the recovered image R will be equal to the original image I, in the case that the difference image (DI) is losslessly compressed. Otherwise, R will deviate from I as dictated by the compression artifacts.

B. Scheme Two

1) Protection Pipeline: The main workflow of our second scheme is similar to the first (Figure 5). However, in this scheme, we capitalize on the coherence between the original image and the inverted false color image to significantly reduce the size of the protected image.

In this scheme, the false color image, FI, is computed somewhat differently to avoid color-to-gray conversion (see Figure 6). For each color value $I_c(x, y)$ in the original image, the corresponding false color value $FI_c(x, y)$ is computed as:

$$FI_c(x, y) = P_c[I_c(x, y)], \tag{6}$$

where P_c denotes the c^{th} channel of the color palette P.

Next, after FI_c is encoded and decoded to obtain FI'_c, instead of directly subtracting it from I, we first apply an inverse table look-up to obtain I':

$$I'_c(x, y) = P_c^{\text{inv}}[FI'_c(x, y)]. \tag{7}$$

Here, P_c^{inv} represents a pseudo-inverse of the c^{th} channel of the color palette. We call it pseudo-inverse as most color palettes are not one-to-one and therefore non-invertible. In practice, given $FI'_c(x, y)$, we search inside P_c to find the index of the most similar color value:

$$I'_c(x, y) = \underset{i \in \{0,1,\ldots,255\}}{\arg\min} |P_c[i] - FI'_c(x, y)|. \tag{8}$$

If there are multiple such values that minimize this difference, we choose the index according to the histogram of the original image. For example, if $i = 5$ and $i = 125$ are two solutions of Equation 8, and $hist(I_c)[125] > hist(I_c)[5]$, we choose 125 as the inverse. This ensures that the inverted value will be similar to the original value for the maximum number of pixels.

Once I'_c is computed in this fashion, it is subtracted from I to obtain the difference and sign images (using Equations 2 and 3 after substituting FI'_c with I'_c). Note that, unlike in Scheme One, the difference image in this case will have many zero or small components that can be compressed efficiently.

Furthermore, an opportunity for a different type of lossy compression presents itself in this particular case. Depending on a quality threshold, τ, all values in DI that are smaller than τ may be set to zero:

$$DI'_c(x, y) = \begin{cases} 0 & \text{if} \quad DI_c(x, y) < \tau, \\ DI_c(x, y) & \text{otherwise.} \end{cases} \tag{9}$$

In this case, the corresponding $SI_c(x, y)$ values should also be set to zero to improve the compression efficiency for the sign image as well. After this process, the difference and sign images are compressed and encrypted before being written to the JPEG application markers. We also compress, encrypt, and write the histograms of each channel and the color palette within the JPEG application markers as well to be used during the recovery process.

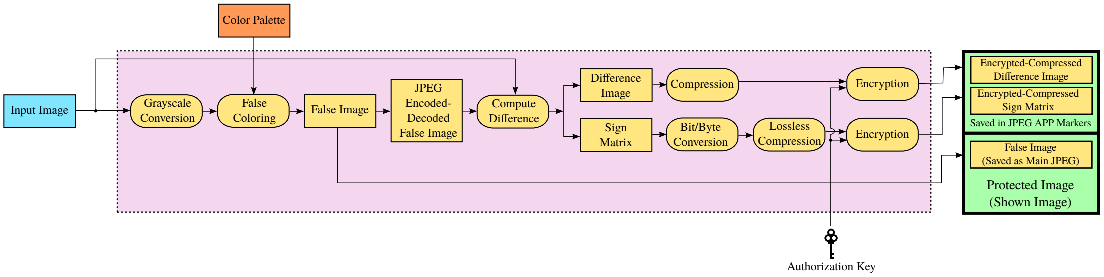


Fig. 3. The protection pipeline for Scheme One.

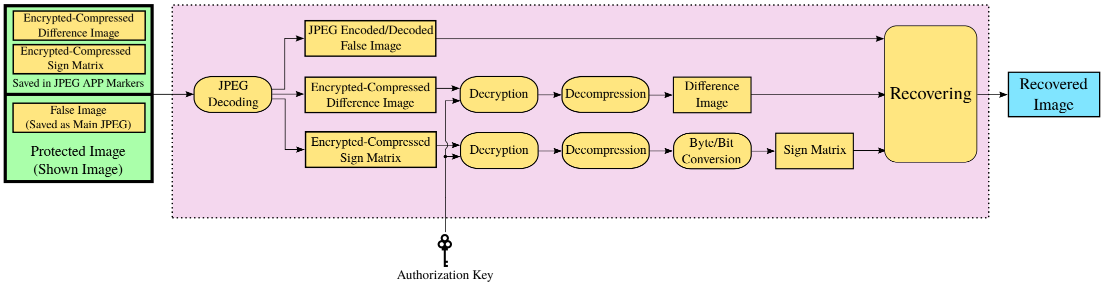

Encrypted-Compressed Difference Image
Encrypted-Compressed Sign Matrix
Saved in JPEG APP Markers
False Image (Saved as Main JPEG)
Protected Image (Shown Image)
JPEG Decoding
JPEG Encoded/Decoded False Image
Encrypted-Compressed Difference Image
Decryption
Decompression
Difference Image
Encrypted-Compressed Sign Matrix
Decryption
Decompression
Byte/Bit Conversion
Sign Matrix
Recovering
Recovered Image
Authorization Key

Fig. 4. The recovery pipeline for Scheme One.

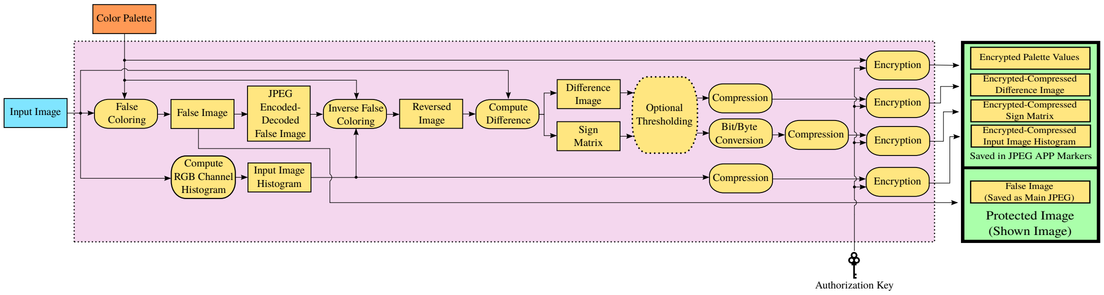
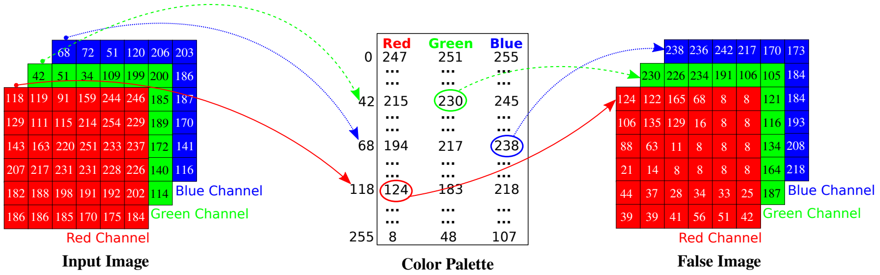


Fig. 5. The protection pipeline for Scheme Two.



Input Image

Blue Channel

68	72	51	120	206	203
					186
					187
					170
					141
					116

Green Channel

42	51	34	109	199	200
					185
					189
					172
					140
					114

Red Channel

118	119	91	159	244	246
129	111	115	214	254	229
143	163	220	251	233	237
207	217	231	231	228	226
182	188	198	191	192	202
186	186	185	170	175	184

Color Palette

	Red	Green	Blue
0	247	251	255
	⋯	⋯	⋯
	⋯	⋯	⋯
42	215	230	245
	⋯	⋯	⋯
	⋯	⋯	⋯
68	194	217	238
	⋯	⋯	⋯
	⋯	⋯	⋯
118	124	183	218
	⋯	⋯	⋯
	⋯	⋯	⋯
255	8	48	107

False Image

Blue Channel

238	236	242	217	170	173
					184
					184
					193
					208
					218

Green Channel

230	226	234	191	106	105
					121
					116
					134
					164
					187

Red Channel

124	122	165	68	8	8
106	135	129	16	8	8
88	63	11	8	8	8
21	14	8	8	8	8
44	37	28	34	33	25
39	39	41	56	51	42

Fig. 6. False coloring method of Scheme Two.

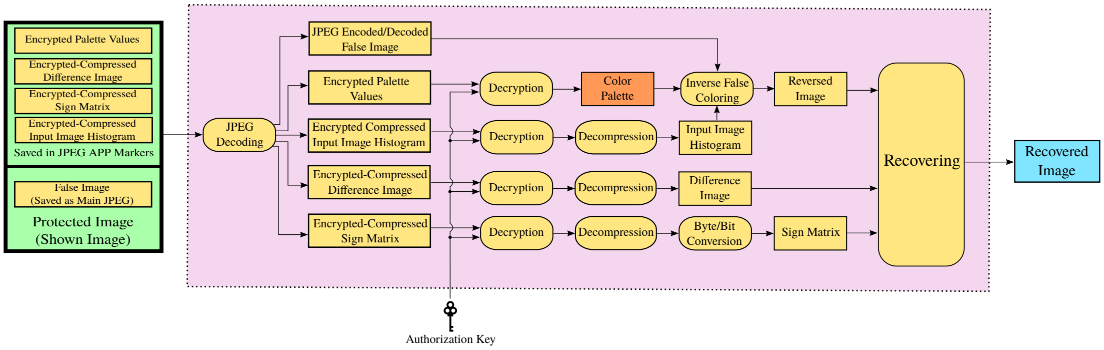
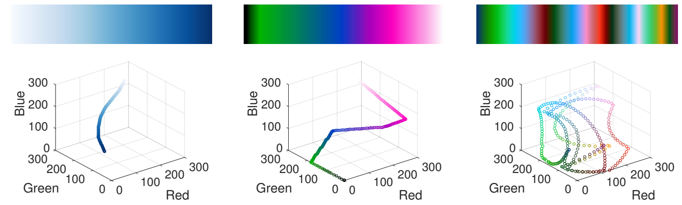

Encrypted Palette Values
Encrypted-Compressed Difference Image
Encrypted-Compressed Sign Matrix
Encrypted-Compressed Input Image Histogram
Saved in JPEG APP Markers
False Image (Saved as Main JPEG)
Protected Image (Shown Image)
JPEG Decoding
JPEG Encoded/Decoded False Image
Encrypted Palette Values
Decryption
Color Palette
Inverse False Coloring
Reversed Image
Encrypted Compressed Input Image Histogram
Decryption
Decompression
Input Image Histogram
Encrypted-Compressed Difference Image
Decryption
Decompression
Difference Image
Encrypted-Compressed Sign Matrix
Decryption
Decompression
Byte/Bit Conversion
Sign Matrix
Recovering
Recovered Image
Authorization Key

Fig. 7. The recovery pipeline for Scheme Two.



a) Blues b) Green-Pink c) Hardcandy

Fig. 8. Characteristics of the color palettes used in this study.

2) Recovery Pipeline: In the recovery pipeline of the second scheme (Figure 7), first all encrypted metadata is decrypted followed by decompression, if needed. Then using the decoded false color image FI', the input image histogram $hist(I_c)$, and the color palette P, an approximate of the original image, I' is computed using Equation 7. This image is then combined with the difference and sign images as in Equation 4 by substituting I'_c for FI'_c. Note that, similar to Scheme One, the recovered image R will be identical to the original image I if the difference and sign images are not thresholded and compressed in a lossless manner.

IV. Results

To help demonstrate the validity of our approach we have conducted a large set of experiments. In this section, we first demonstrate the visual quality of our outputs in lossless protection mode for both schemes. We then illustrate the effect of lossy compression on both file size and recovered image quality. This is followed by a comparison with other well-known privacy protection approaches using a face recognition benchmark. Next, we share the results of subjective and objective evaluations conducted on a novel dataset, in which we compare the privacy vs. intelligibility trade-off of our algorithm with commonly used VPP approaches. Finally, we investigate whether a simulated attack on false color protected content could effectively reverse protection compromising the privacy of the recorded individuals.

In our experiments, we first analyzed a large number of color palettes available in National Library of Medicine Insight Segmentation and Registration Toolkit (ITK) [49]. We decided to use 3 color palettes based on their apparent effectiveness in preserving privacy. Among these, the *Blues* palette has a more monotonic variation of colors, *Hardcandy* is extremely erratic, and the *Green-Pink* is in-between. These palettes are shown in Figure 8 together with 3D scatter plots that show the path traversed by each palette within the RGB color space.

A. Lossless Compression

We first illustrate our results for the *Protest* scene shown in Figure 9. In this scene, 6 people comprised of 3 males and 3 females are holding banners and simulating a protestation scenario. Each row represents a different color palette. The first column shows the privacy protected false color images, the second column shows the difference images and the third column shows the sign images (same for Scheme Two in the last three columns). In the sign images, negative differences are indicated by 1 (mapped to 255 for illustration purposes) and positive differences by 0. Because this difference is computed for each channel, it contains colors made up of combination of the three primary colors. For all three color palettes and for both schemes, it can be observed that while the protected images are still intelligible, the identities of the people are mostly concealed.

By inspecting these figures, one can observe the differences between the two schemes as well as the effect of the color palette. First, it can be observed that the false color images are very similar for both schemes. As for the difference images, the second scheme yields images that have smaller values than in Scheme One. This is due to the recovery step (Equations 7 and 8) applied in the second scheme. However, the color palettes also influences the quality of the recovery. For the *Blues* palette, which is more monotonic than the other two palettes (Figure 8), the quality of the recovery is very well and therefore the difference image contains very small values. However, for the other two palettes the recovery is progressively less effective due to their less regular variations across the color scale.

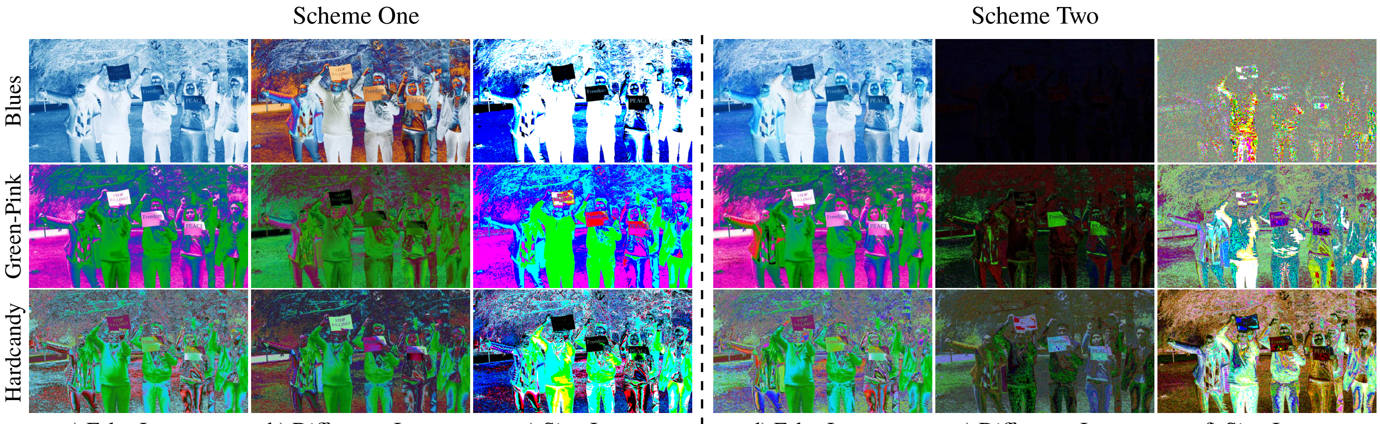
As for the sign images, they appear to be more noisy as the difference images have smaller pixel values. This is because the small differences may be positive or negative, and this may change rapidly from pixel to pixel (even between color channels of a pixel).

These observations are also supported by the byte-sizes of these components in both schemes (Table I). Here, FI represents the JPEG-compressed size of the false color image at JPEG quality setting 85; DI and SI represent the lossless zlib-compressed size of the difference and sign images; and PI represent the total bitstream size of the protected images. As shown in this table, the DI in Scheme Two is smaller than that of Scheme One, whereas the SI is larger. Also, the DI value in Scheme Two increases with the variance of the palette.

B. Lossy Compression

In many applications, the file size of the protected images may be critical and the extra overhead introduced by the required metadata may be too large. For this purpose, we also propose a lossy mode and present our experimental results in terms of image quality versus file size.

As the chief overhead is introduced by the difference image, we explored several techniques to reduce its size. In Scheme One, we experimented with storing it with JPEG compression as well as by downsampling it. Our results are shown in Figure 10. In the first three columns of this figure, the JPEG quality value of the difference image, Q_D, is varied as 85, 50, and 10. In the last three columns, the difference image is downsampled with factors of 2, 8, and 16. The images shown are the recovered images. Structural similarity index (SSIM) values [50] with respect to the original images are also reported. As expected, the visual quality of the recovered images degrades with the increased compression of the difference image. However, the file size of the protected images also gets smaller. Of the two compression techniques, JPEG encoding the difference image appears to be a better approach



a) False Image b) Difference Image c) Sign Image d) False Image e) Difference Image f) Sign Image

Fig. 9. Visual representation of false color based protection components for Scheme One and Scheme Two.

TABLE I
THE FILE SIZE OF VARIOUS COMPONENTS IN BOTH SCHEMES. THE ORIGINAL IMAGE HAS A RESOLUTION OF 1920 × 1080 AND OCCUPIES 674.36 KBS OF DISK SPACE.

	Scheme One				**Scheme Two**			
	FI	**DI**	**SI**	**PI**	**FI**	**DI**	**SI**	**PI**
Blues	486.01KB	5.23MB	183.25KB	5.89MB	492.79KB	3.21MB	634.77KB	4.31MB
Green-Pink	596.61KB	5.06MB	173.83KB	5.82MB	624.01KB	4.22MB	614.28KB	5.43MB
Hardcandy	1.02MB	5.34MB	438.05KB	6.79MB	1.02MB	5.18MB	638.86KB	6.83MB

as it not only produces smaller files but also better maintains the visual quality.

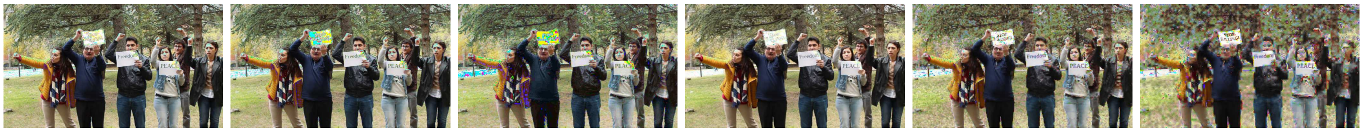
While the same compression strategies can be used for Scheme Two as well, a different type of compression strategy can be employed. As the difference image in Scheme Two is computed after performing an inverse color-palette look-up, it contains many small values. By defining a threshold parameter, τ, below which these differences are set to zero, one can further make the difference image more compressible. Our results for the effectiveness of this approach are shown in Figure 11. In the first three images we show that for the *Blues* color palette as the τ value is increased, the loss in the recovered image quality remains almost negligible despite a significant reduction in file size. In the last three images, the same is shown for the less regular *Hardcandy* color palette. As can be seen by the SSIM score and the file size, this approach is less effective for this color palette compared to the *Blues* palette, as in the latter the difference image does not contain many small values.

To allow generalization of our results, we captured several videos that simulate various surveillance scenarios. From each video, we selected a representative frame resulting in a total of 12 test images. We then performed the lossy compression approaches described earlier. Our results are reported in Table II. For the lossless case, the best compression ratio is provided by the second scheme with the *Blues* palette (7.07). For the JPEG encoding vs. downsampling of the first scheme, the former yields not only better compression but also higher average SSIM scores. By increasing the τ threshold in the second scheme, one can obtain protected files that are even smaller than the original files. However, for highly irregular color palettes, such as *Hardcandy*, even high threshold values do not produce very small files as the difference image contains many pixels that are above this threshold.

C. *Face recognition*

In this section, we illustrate the performance of our algorithm and compare it with other privacy protection techniques using a face recognition benchmarking framework developed by Korshunov et al. [51]. This framework uses three different face recognition algorithms (FRAs), namely Eigenfaces [52], Fisherfaces [53], and LBPH [54], which are all implemented in OpenCV. As for the dataset, it uses the FERET face image dataset that contains multiple face images of 100 individuals [55]. For comparison, we used the three filters that are provided with this framework. These are blurring, pixelation, and warping [23]. Sample outputs of our method and the compared filters are shown in Figure 12.

The aggregated results are shown in Table III. In this table, the lower the value, the better the protection performance. For all three FRAs and all three color palettes, the performance of false color based privacy is very good with the highest recognition rate being 0.14 for the *LBPH-Green-Pink* combination. The *Blues* and *Hardcandy* palettes perform very well and their recognition rates are not higher than the chance level of 1% for the Eigenfaces and Fisherfaces algorithms. Of the compared algorithms, our method is outperformed only by pixelation with the LBPH algorithm. However, as shown in Figure 12, a pixelation window size of 5 is not effective at preserving privacy. On the other end, a window size of 55 prevents intelligibility and reversibility entirely. Furthermore, the performance of pixelation is not good under other FRAs.



Q_D: 85, SSIM: 0.91 PI: 1.32 MB

Q_D: 50, SSIM: 0.82 PI: 1018.92 KB

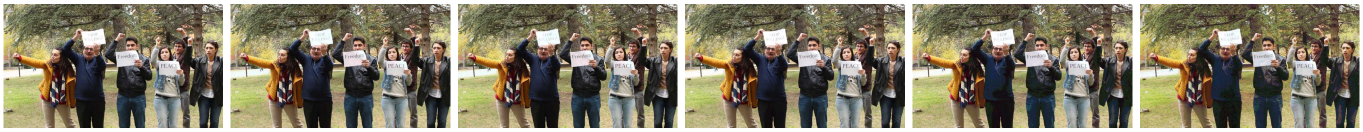
Q_D: 10, SSIM: 0.62 PI: 796.57 KB

DS: 2, SSIM: 0.75 PI: 2.02 MB

DS: 8, SSIM: 0.29 PI: 761.17 KB

DS: 16, SSIM: 0.22 PI: 692.59 KB

Fig. 10. Recovery results for lossy compressions in Scheme One. Q_D, DS, and PI namely stands for JPEG compression quality of the difference image, down sampling factor of the same, and protected image file size. Results are shown for the *Blues* color palette.



τ: 10, SSIM: 0.99 PI: 2.16 MB

τ: 40, SSIM: 0.97 PI: 732.71 KB

τ: 70, SSIM: 0.97 PI: 516.53 KB

τ: 10, SSIM: 0.99 PI: 6.40 MB

τ: 40, SSIM: 0.86 PI: 5.41 MB

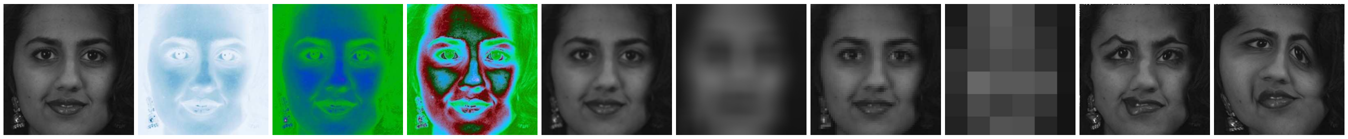
τ: 70, SSIM: 0.64 PI: 4.53 MB

Fig. 11. Recovery results for lossy compressions in Scheme Two. τ stands for thresholding value for the difference image. Results are shown for Blues (left-three) and Hardcandy (right-three) color palettes.

TABLE II

PROTECTED TO ORIGINAL IMAGE FILE SIZE RATIOS AGGREGATED OVER 12 TEST IMAGES. SSIM SCORES ARE SHOWN AFTER THE COMMA. THE FALSE COLOR IMAGE JPEG ENCODING QUALITY IS FIXED AT 85.

	Scheme One							**Scheme Two**			
	Lossless	Q_D**: 85**	Q_D**: 50**	Q_D**: 10**	DS**: 2**	DS**: 8**	DS**: 16**	**Lossless**	τ**: 10**	τ**: 40**	τ**: 70**
Blues	9.95	1.91, 0.92	1.44, 0.84	1.12, 0.65	3.40, 0.78	1.11, 0.39	0.98, 0.33	7.07	2.72, 0.99	0.94, 0.98	0.74, 0.97
Green-Pink	10.08	1.92, 0.89	1.49, 0.82	1.25, 0.71	3.48, 0.87	1.27, 0.64	1.16, 0.58	9.37	6.70, 0.98	4.12, 0.86	2.67, 0.71
Hardcandy	12.09	3.47, 0.82	2.76, 0.68	2.35, 0.53	4.69, 0.69	2.31, 0.46	2.19, 0.41	12.25	11.35, 0.99	9.45, 0.82	7.90, 0.56



Original Image

Blues

Green-Pink

Hardcandy

Blurring (5)

Blurring (55)

Pixelation (5)

Pixelation (55)

Warping (13)

Warping (3)

Fig. 12. Sample outputs of the compared methods for a face image from FERET dataset. Numbers in parenthesis indicate the blurring kernel size, pixelation window size, and warping strength value (note that this value is inversely proportional to the degree of warping). Images' resolution was 320 × 320.

Based on these results, it can be argued that false color based privacy protection outperforms the compared approaches on this evaluation task.

We also used the same evaluation framework to assess the recovery performance of our algorithm. To this end, protected images with various degrees of lossy compression are recovered to obtain approximations of the original images. The overall face recognition results on the recovered face images are reported in Table IV. The recovery results with respect to Eigenfaces and Fisherfaces are all around 90%. This is very close to the baseline performance that would be obtained if the recovered images were identical to the originals. For the LBPH algorithm, the *Blues* palette performs the best, followed by the *Green-Pink* and *Hardcandy* palettes.

To summarize, the face recognition evaluation indicates that the proposed false coloring algorithm protects privacy better than the compared approaches. Furthermore, the protected images can be reversed to obtain recovered images which can be recognized by FRAs, even when compressed by lossy compression algorithms.

D. Surveillance Video Dataset

We created a novel surveillance video dataset, called METUSURV[1], in order to better understand the intelligibility vs. privacy trade-off of the proposed algorithm and compare it with other VPP algorithms. This dataset includes several security related scenarios such as fighting, protesting, stealing, physical/verbal harassment, bag leaving, and bag exchanging. While there are such datasets in the literature, the distinguishing feature of METUSURV is that it also contains face images of the people that are recorded in these videos. The face images were captured in a controlled room and each face was represented using three images, one frontal and two profile. In total, 60 individuals' (40M and 20F) faces were captured. A subset of these individuals acted in the recorded videos.

As for the recording, various surveillance situations were enacted. Some of the people in these videos were not part of the face image dataset. Of a larger number of captured videos, 12 of them were found to have sufficient quality to be included in the dataset. All of these videos were cropped to

[1] Authors may be contacted to access and use this dataset for research purposes.

TABLE III
FACE RECOGNITION ACCURACY RATES. THE LOWER THE VALUE, THE BETTER THE PROTECTION PERFORMANCE. STR. LEV., KER. SIZE, WIN. SIZE RESPECTIVELY DENOTES WARPING STRENGTH LEVEL, BLURRING KERNEL SIZE, AND PIXELATION WINDOW SIZE. AS THE WARPING ALGORITHM RANDOMLY DETERMINES THE INITIAL WARPING POINTS, IT IS RUN FOR 10 ITERATIONS AND THE MEAN ACCURACY VALUES ARE REPORTED.

<table>
<tr><th></th><th colspan="6">LBPH</th><th colspan="6">Eigen</th><th colspan="6">Fisher</th></tr>
<tr><td rowspan="2"><b>Warping</b></td><td colspan="3"><b>Str. Lev.: 3</b></td><td colspan="3"><b>Str. Lev.: 13</b></td><td colspan="3"><b>Str. Lev.: 3</b></td><td colspan="3"><b>Str. Lev.: 13</b></td><td colspan="3"><b>Str. Lev.: 3</b></td><td colspan="3"><b>Str. Lev.: 13</b></td></tr>
<tr><td colspan="3">0.75</td><td colspan="3">0.91</td><td colspan="3">0.77</td><td colspan="3">0.89</td><td colspan="3">0.76</td><td colspan="3">0.89</td></tr>
<tr><td rowspan="2"><b>Blurring</b></td><td colspan="3"><b>Ker. Size: 5</b></td><td colspan="3"><b>Ker. Size: 55</b></td><td colspan="3"><b>Ker. Size: 5</b></td><td colspan="3"><b>Ker. Size: 55</b></td><td colspan="3"><b>Ker. Size: 5</b></td><td colspan="3"><b>Ker. Size: 55</b></td></tr>
<tr><td colspan="3">0.72</td><td colspan="3">0.14</td><td colspan="3">0.89</td><td colspan="3">0.79</td><td colspan="3">0.89</td><td colspan="3">0.79</td></tr>
<tr><td rowspan="2"><b>Pixelation</b></td><td colspan="3"><b>Win. Size: 5</b></td><td colspan="3"><b>Win. Size: 55</b></td><td colspan="3"><b>Win. Size: 5</b></td><td colspan="3"><b>Win. Size: 55</b></td><td colspan="3"><b>Win. Size: 5</b></td><td colspan="3"><b>Win. Size: 55</b></td></tr>
<tr><td colspan="3">0.04</td><td colspan="3">0.00</td><td colspan="3">0.89</td><td colspan="3">0.55</td><td colspan="3">0.89</td><td colspan="3">0.55</td></tr>
<tr><td rowspan="2"><b>False Coloring</b></td><td colspan="2"><b>Blues</b></td><td colspan="2"><b>Green-Pink</b></td><td colspan="2"><b>Hardcandy</b></td><td colspan="2"><b>Blues</b></td><td colspan="2"><b>Green-Pink</b></td><td colspan="2"><b>Hardcandy</b></td><td colspan="2"><b>Blues</b></td><td colspan="2"><b>Green-Pink</b></td><td colspan="2"><b>Hardcandy</b></td></tr>
<tr><td colspan="2">0.07</td><td colspan="2">0.14</td><td colspan="2">0.09</td><td colspan="2">0.00</td><td colspan="2">0.07</td><td colspan="2">0.01</td><td colspan="2">0.00</td><td colspan="2">0.07</td><td colspan="2">0.01</td></tr>
</table>

TABLE IV
FACE RECOGNITION ACCURACY RATES FOR RECOVERED IMAGES. THE CLOSER TO 1, THE BETTER THE ACCURACY.

	LBPH									**Eigen-Fisher**								
	Q_D			DS			τ			Q_D			DS			τ		
Param. / **Pal.**	**85**	**50**	**10**	**2**	**8**	**16**	**10**	**40**	**70**	**85**	**50**	**10**	**2**	**8**	**16**	**10**	**40**	**70**
Blues	0.91	0.83	0.81	0.90	0.88	0.78	0.87	0.89	0.88	0.89	0.88	0.88	0.89	0.89	0.88	0.89	0.89	0.89
Green-Pink	0.91	0.86	0.55	0.88	0.86	0.79	0.84	0.77	0.69	0.89	0.89	0.89	0.89	0.89	0.87	0.89	0.89	0.89
Hardcandy	0.34	0.07	0.03	0.15	0.10	0.06	0.90	0.07	0.02	0.89	0.89	0.88	0.89	0.86	0.82-0.83	0.89	0.89	0.89

approximately 15 seconds. All videos and face images were captured using a Canon EOS 600D DSLR camera.

Furthermore, each surveillance scenario was recorded several times using the same group of actors. One of these recordings was selected as the main video and the others were used to train the face recognition algorithms as discussed in Section IV-D2.

1) Subjective Evaluation: We conducted a user study to evaluate the effectiveness of false coloring on privacy and intelligibility. In our evaluation, 6 videos were selected from the METUSURV dataset. For false coloring, the *Blues* palette was selected as representative of our method as the other color palettes produced a similar result in the face recognition benchmark. For comparison, we opted for blurring and pixelation as they are commonly used and can also be applied globally while maintaining intelligibility. Blurring was performed using a Gaussian kernel with $\sigma = 9$ pixels evaluated over a window size of 55×55. The block size for pixelation was selected as 10×10.

Through a web-based interface, we asked each participant to indicate which people were present in a given video and whether they noticed any suspicious activities. The experimental interface is shown in Figure 13.

The participants were instructed to first study the six faces shown in order to identify them in the upcoming video. Exactly half of these faces were present resulting in a chance estimation rate of 50%. Once the participants were ready, they clicked the “Play Video for Face Recognition” button to view the privacy protected video. The video could be seen only once. After the video was shown, the participants were shown the face pictures again to collect their responses.

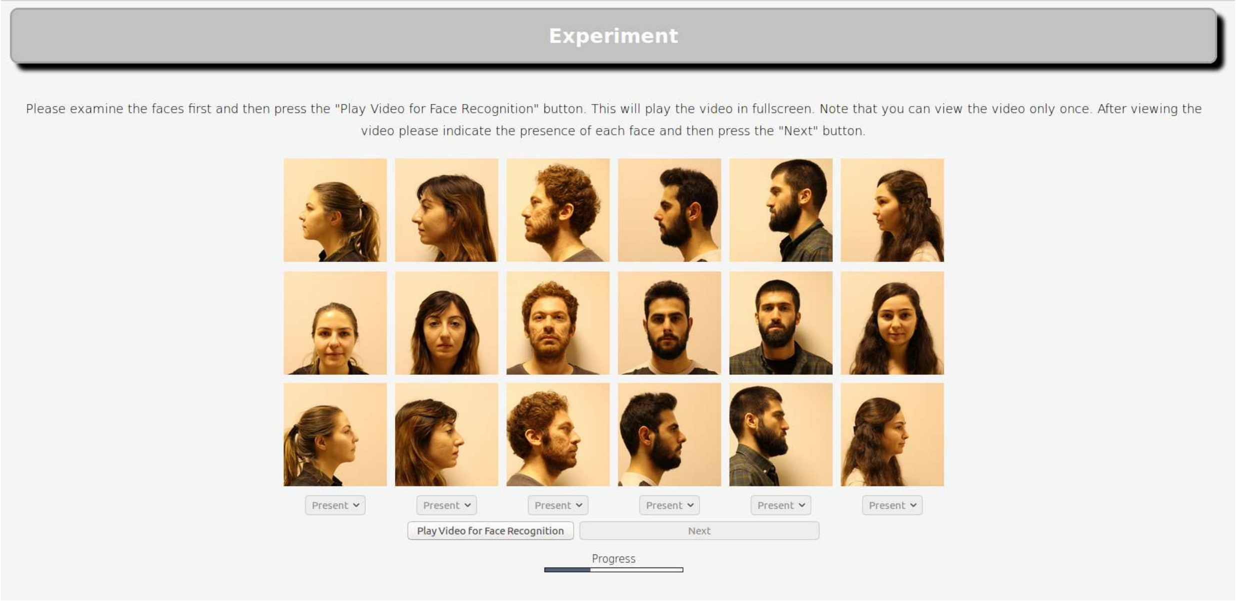
Next, the participants clicked the “Play Video for Activity Analysis” button, which resulted in the presentation of the same video for the second time. This time the participants' task was to identify the potentially suspicious activities that were taking place in the video. Again, the video was shown only once and when it ended the participants were shown a page in which they could select the activities they observed (Figure 14). The reason for separating the face recognition task from the activity analysis task was that coping with both of them simultaneously proved to be very challenging during the pilot runs of the experiment.

Each participant viewed 6 different videos (2 videos for each of the 3 VPP algorithms) throughout the experiment. While the presentation order of the videos was random, the experiment was designed so that the same video was not shown twice with a different privacy protection method. This was done to eliminate a potential memory effect between the methods. Also, the pairing between the face images and the videos was done manually to ensure that for each trial half of the 6 faces were present in the video. The display order of the faces on the face recognition screen (Figure 13) was random.

In total, 48 participants (19 females, 29 males) participated in the experiment. None of the participants were acquainted with the people shown in the videos to eliminate other cues from affecting their decisions. Each participant could finish the experiment within 15 minutes.

The results of the experiment are given in Tables V and VI. In Table V, it can be seen that the mean face recognition rate for false coloring, 0.57, is the lowest among the compared methods (0.59 for blurring and 0.62 for pixelation). Note that this rate is close to the chance rate of 0.50, that would be achieved if the participants were making entirely random decisions. As for the activity detection, Table VI shows the f-scores corresponding to each algorithm. Here, it can be seen that participants could better observe the suspicious activities under false coloring (0.86) than under blurring (0.84) and pixelation (0.84).

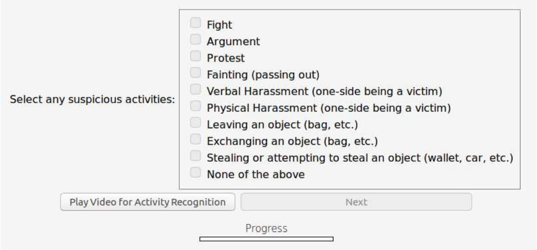
To summarize, the subjective evaluation results suggest that false coloring preserves privacy better than the compared methods while having a less impact on intelligibility. We, however, note that this evaluation only considered the *who* and *what* questions that pertain to privacy and intelligibility. As



Experiment

Please examine the faces first and then press the "Play Video for Face Recognition" button. This will play the video in fullscreen. Note that you can view the video only once. After viewing the video please indicate the presence of each face and then press the "Next" button.

Fig. 13. The experimental interface for face recognition.



Select any suspicious activities:

- □ Fight
- □ Argument
- □ Protest
- □ Fainting (passing out)
- □ Verbal Harassment (one-side being a victim)
- □ Physical Harassment (one-side being a victim)
- □ Leaving an object (bag, etc.)
- □ Exchanging an object (bag, etc.)
- □ Stealing or attempting to steal an object (wallet, car, etc.)
- □ None of the above

Fig. 14. The experimental interface for activity recognition.

suggested by Saini et al. [47], the questions of *where* and *when* may also have an effect on both attributes. For instance, if one can infer where a footage was captured, when it was captured, and what actions were taking place, he/she can determine who the people are without explicitly recognizing their faces. The design of our experiment, especially the requirement that the participants did not know about the actors, was not suitable for this type of analysis. Furthermore, as in any evaluation, the choice of the parameters might have affected the specific outcomes obtained. However, due to the difficulty of running a subjective experiment over a large number of parameter values, a sensible set of values were determined by visually inspecting the outputs to obtain a balance between privacy and intelligibility.

2) Objective Evaluation: Ensuing the subjective evaluation, objective experiments were also conducted to understand how well face recognition algorithms (FRAs) can successfully recognize the people in the protected videos.

The FRAs need a training image set for creating a model that is used for identifying the detected faces in a given image. Face images in the METUSURV dataset were taken in an indoor environment and they were not sufficient to train a model for recognizing people in the videos that were captured in an outdoor environment. For that reason, a new training dataset was created by first executing the Viola-Jones face detection algorithm [17] on each frame of the captured surveillance videos. Each surveillance scenario was recorded several times using the same group of actors. One of these recordings was selected as the main video (this was the video shown to the participants in the subjective evaluation) and the others were used to train the face recognition algorithms.

The regions returned by the face detection algorithm were visually observed to ensure that they correspond to faces. Furthermore, some of the faces could not be found by the algorithm as they were distorted by motion-blur artifacts. These faces were manually cropped and put into the training set. For each face, at least 10 different training images were thus selected. Eigenfaces [52], Fisherfaces [53], and LBPH [54] were used as the face recognition algorithms.

The generated face recognition models were tested on videos that were protected with blurring, pixelation, and false coloring VPP algorithms. The evaluation of the results were performed with respect to three different definitions of *true positive* (TP) and *false positive* (FP).

In Evaluation 1, the TP count was incremented whenever a region (returned by the face detector) was identified as a person that was present in the video. The FP count was incremented when an identified person was not present in the video. In this evaluation, it is important to note that the region returned by the face detector may not actually correspond to a face region. In Evaluation 2, non-face regions returned by the face detector were manually removed from testing. Finally, in Evaluation 3, in addition to removing non-face regions, the TP count was only incremented if a face region was attributed to the correct person. Likewise, the FP count was only incremented when a face region was attributed to a wrong person. To this end, Evaluation 3 can be considered as the most refined measure as it considers correctness of each

TABLE V
FACE RECOGNITION RATES OBTAINED IN THE SUBJECTIVE EVALUATION AVERAGED OVER ALL PARTICIPANTS. THE RANGE OF ALL SCORES IS IN $[0, 1]$.

	Video 1	**Video 2**	**Video 3**	**Video 4**	**Video 5**	**Video 6**	**Average**
Blurring	0.73	0.47	0.56	0.60	0.65	0.53	0.59
Pixelation	0.78	0.61	0.55	0.58	0.69	0.54	0.62
False Coloring	0.64	0.43	0.54	0.55	0.69	0.60	**0.57**

TABLE VI
SUSPICIOUS ACTIVITY DETECTION RESULTS.

	Video 1	**Video 2**	**Video 3**	**Video 4**	**Video 5**	**Video 6**	**Average**
Blurring	0.81	0.89	0.74	1.00	0.71	0.90	0.84
Pixelation	0.78	0.91	0.74	0.97	0.71	0.92	0.84
False Coloring	0.79	0.93	0.76	0.98	0.75	0.91	**0.86**

region individually.

As for *true negative* (TN) and *false negative* (FN), all evaluations used the same criteria. The TN count was incremented whenever a person not present in the video (but present in the training set) was not identified. The FN count was incremented when a person present in the video was not identified. The evaluations were made with respect to precision, recall, and f-score. The overall results computed by taking the average of the 6 videos are shown in Table VII.

In this table, it can be seen that false coloring produces the lowest f-score in all evaluations except in Evaluation 1/Fisherfaces combination. For that combination, blurring yields a slightly lower score than false coloring. The *unfiltered* row in this table shows the precision, recall, and f-score results for unprotected videos. It is important to note that the precision for the unfiltered videos increases from Evaluation 1 to Evaluation 2. This is because as non-face regions are removed, the FRAs are less likely to make Type-1 error. However, the same pattern is not observed especially for false coloring results. This is because as non-face regions are removed, in many cases, the remaining number of regions on which FRAs operate become zero or very limited. This suggests that often recognition of a face in a false colored video fails in the detection step. Furthermore, it is important to note that the precision obtained in Evaluation 3 is consistently lower than the precision in Evaluation 2 for all presentation types. This is expected because the success criteria of Evaluation 3 was more strict as explained above.

E. Security Evaluation

Besides privacy and intelligibility, security against unauthorized users is another desirable property of an effective VPP algorithm. Here, we evaluate whether false coloring is secure against an attacker who tries to revert the protected face images to their unprotected versions. To this end, we selected 12 face images from the FERET dataset representing people of different ethnicities (Figure 15). From each face, we selected pixels that correspond to skin, lip, and hair regions. For each region, pixel values within a 5×5 neighborhood were averaged to obtain a representative color.

Fig. 15. Images selected from the FERET dataset for security evaluation.

Next, a target face transformed by our second scheme was selected. The false color RGB value for the corresponding face regions were computed giving rise to the following mapping:

$$G_r \rightarrow RGB_r. \qquad (10)$$

Here, G_r represents the grayscale value of region $r \in \{\mathrm{skin}, \mathrm{hair}, \mathrm{lip}\}$ from the FERET image and RGB_r represents the corresponding false color value from the target face. This mapping was then sorted by increasing G_r value, and the missing values of the color palette were computed by linear interpolation. Note that as we used 12 input images, this resulted in 12 reconstructed palettes for each target face.

Next, an inverse look-up to each reconstructed color palette was produced to find the corresponding grayscale value for each false color pixel. Some palettes produced better results than others and we visually determined the best palette as the most accurate reversal of a target face. The results are shown in Figure 16 for each of the false color palettes evaluated in this study. As can be seen in this figure, while the *Blues* and *Green-Pink* palettes are vulnerable to this attack, the *Hardcandy* palette remains resistant for each target face. This evaluation leads us to conclude that if security is the primary concern, one should select a palette with more random variation to avoid reconstruction of the original data by unauthorized users.

V. CONCLUSIONS & FUTURE WORK

We presented a false color based privacy protection algorithm implemented within the JPEG architecture and demonstrated its performance by conducting extensive experiments. In particular, we have shown our method to be effective against not only face recognition algorithms but also against human observers through objective and subjective evaluations.

The primary advantages of our method is that it can be applied on entire images, obviating the need to define privacy sensitive ROIs. This is important because while manually

TABLE VII
FACE RECOGNITION ACCURACY RESULTS ON PROTECTED VIDEOS. LOWER VALUES INDICATE BETTER PROTECTION. PREC. AND REC. INDICATES PRECISION AND RECALL VALUES.

		Evaluation 1			**Evaluation 2**			**Evaluation 3**		
		Prec.	**Rec.**	**F-Score**	**Prec.**	**Rec.**	**F-Score**	**Prec.**	**Rec.**	**F-score**
Eigenfaces	**Blurring**	0.27	0.32	0.28	0.25	0.27	0.24	0.12	0.17	0.13
	Pixelation	0.35	0.47	0.36	0.49	0.47	0.39	0.41	0.40	0.32
	False Coloring	0.28	0.08	**0.12**	0.17	0.02	**0.04**	0.00	0.00	**0.00**
	Unfiltered	0.48	0.83	0.56	0.51	0.83	0.58	0.42	0.68	0.48
Fisherfaces	**Blurring**	0.30	0.26	**0.25**	0.16	0.20	0.17	0.15	0.17	0.15
	Pixelation	0.27	0.51	0.30	0.27	0.47	0.29	0.15	0.34	0.19
	False Coloring	0.29	0.30	0.27	0.08	0.02	**0.03**	0.00	0.00	**0.00**
	Unfiltered	0.33	0.78	0.42	0.34	0.78	0.42	0.28	0.66	0.35
LBPH	**Blurring**	0.24	0.44	0.29	0.14	0.32	0.20	0.14	0.32	0.20
	Pixelation	0.21	0.39	0.26	0.19	0.35	0.24	0.15	0.25	0.17
	False Coloring	0.20	0.16	**0.17**	0.06	0.02	**0.03**	0.00	0.00	**0.00**
	Unfiltered	0.55	0.78	0.61	0.64	0.78	0.65	0.58	0.71	0.59

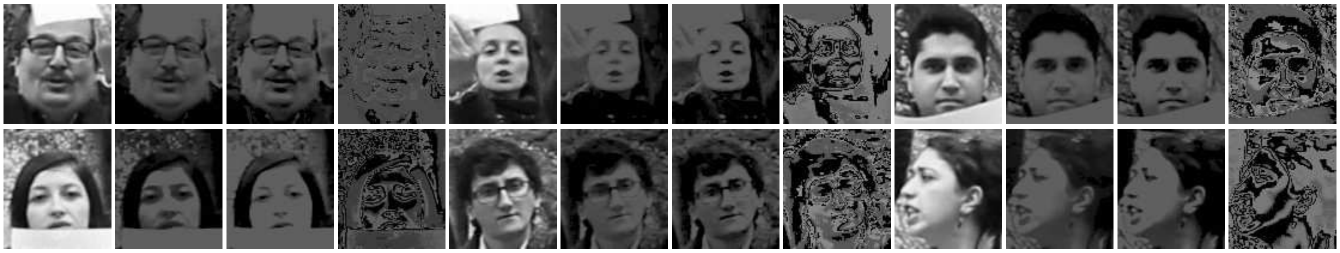


Fig. 16. The best reconstruction results by inverting a color palette using FERET images (see text for details).

defining ROIs is cumbersome and its utility is limited to static scenes, automatic selection of ROIs through detection algorithms is subject to robustness of these algorithms.

The selection of a suitable color palette was found to be an important aspect of our method. Regular (i.e. monotonically varying) color palettes which exhibit inverse relationship between luminance and color saturation (i.e. low luminances are represented with more saturated colors and high luminances are represented with less saturated ones) are found to be effective in preserving privacy, a finding supported by earlier work [56]. Furthermore, regular color palettes result in very small protected file sizes, especially if minor losses are tolerable as afforded by the τ parameter in the second scheme.

As for security, however, regular color palettes are more likely to be decipherable by unauthorized individuals by mapping out relationships between false color pixel values and real object colors as demonstrated by our security evaluation experiments. Irregular palettes, such as *Hardcandy*, are found to provide higher security. However, there is a balance between security and intelligibility: a completely random palette would be very secure but not intelligible as all structural details would be lost. Also, as shown in the previous section, using more irregular palettes results in larger protected file sizes. Perhaps, the most desirable approach would be to define custom color palettes specifically designed for privacy protection purposes – an issue that we leave for future work.

We note that the dependence of the results on the color palette is not necessarily a weakness of our method, but it rather highlights the fact that suitable palettes should be chosen by considering the goals of the application. If the goal is to maximize privacy and security, irregular palettes should be preferred. If the goal is to provide a certain degree of privacy, more regular color palettes could be more suitable. As such, we consider the color palette as a parameter of our method with the best one to be decided based on the actual use case.

It is important to note that the idea of storing the difference and sign images as metadata may apply to other VPP approaches such as blurring and pixelation. However, in false coloring storing the color palette and histogram allows for a greater reduction in file size compared to the other approaches. This is in addition to the fact that false coloring outperforms both approaches with respect to an extensive set of evaluations as conducted in this study.

Finally, although the proposed solution is illustrated for JPEG images, it is possible to apply it in a simple manner to any other image and video formats that support inclusion of metadata. For instance, in AVC/H.264 and HEVC/H.265 formats, this mechanism can be implemented via Supplemental Enhancement Information (SEI) markers. Such an extension is likely to facilitate more widespread adoption of false color based privacy protection by real-world surveillance systems.

VI. ACKNOWLEDGMENT

This work was supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK) project 114E445.

REFERENCES

[1] B. S. I. Association, “The picture is not clear: How many CCTV surveillance cameras in the UK?” July 2013, form No. 195, Issue 1.0.

[2] S. Fleck and W. Straßer, “Smart camera based monitoring system and its application to assisted living,” *Proc. of the IEEE*, vol. 96, no. 10, pp. 1698–1714, 2008.

[3] X. Yuan, X. Wang, C. Wang, J. Weng, and K. Ren, “Enabling secure and fast indexing for privacy-assured healthcare monitoring via compressive sensing,” *IEEE Trans. on Multimedia*, vol. 18, no. 10, pp. 2002–2014, Oct 2016.

[4] D. A. Rodríguez-Silva, L. Adkinson-Orellana, F. Gonz'lez-Castano, I. Armino-Franco, and D. Gonz'lez-Martinez, "Video surveillance based on cloud storage," in *Cloud Computing (CLOUD), 2012 IEEE 5th Intl. Conf. on*. IEEE, 2012, pp. 991–992.

[5] Z. Qian, X. Zhang, and S. Wang, "Reversible data hiding in encrypted jpeg bitstream," *IEEE Trans. on Multimedia*, vol. 16, no. 5, pp. 1486–1491, Aug 2014.

[6] W. Zhang, H. Wang, D. Hou, and N. Yu, "Reversible data hiding in encrypted images by reversible image transformation," *IEEE Trans. on Multimedia*, vol. 18, no. 8, pp. 1469–1479, Aug 2016.

[7] J. R. Padilla-López, A. A. Chaaraoui, and F. Flórez-Revuelta, "Visual privacy protection methods: A survey," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4177–4195, 2015.

[8] S. Ribaric, A. Ariyaeeinia, and N. Pavesic, "De-identification for privacy protection in multimedia content: A survey," *Signal Processing: Image Communication*, vol. 47, pp. 131–151, 2016.

[9] K. Lander, V. Bruce, and H. Hill, "Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces," *Applied Cog. Psyc.*, vol. 15, no. 1, pp. 101–116, 2001.

[10] E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *IEEE Trans. on Know. and Data Engr.*, vol. 17, no. 2, pp. 232–243, 2005.

[11] B. Furht, D. Socek, and A. M. Eskicioglu, "Fundamentals of multimedia encryption techniques," *Multimedia Security Handbook*, vol. 4, 2004.

[12] W. Zeng and S. Lei, "Efficient frequency domain selective scrambling of digital video," *Multimedia, IEEE Trans. on*, vol. 5, no. 1, pp. 118–129, 2003.

[13] L. Tang, "Methods for encrypting and decrypting mpeg video data efficiently," in *Proc. of the Fourth ACM Intl. Conf. on Multimedia*. ACM, 1997, pp. 219–229.

[14] F. Dufaux and T. Ebrahimi, "Scrambling for privacy protection in video surveillance systems," *Circuits and systems for video technology, IEEE Trans. on*, vol. 18, no. 8, pp. 1168–1174, 2008.

[15] ——, "A framework for the validation of privacy protection solutions in video surveillance," in *Multimedia and Expo (ICME), 2010 IEEE Intl. Conf. on*. IEEE, 2010, pp. 66–71.

[16] F. Dufaux, "Video scrambling for privacy protection in video surveillance: recent results and validation framework," in *SPIE Defense, Security, and Sensing*. Intl. Society for Optics and Photonics, 2011, pp. 806 302–806 302.

[17] P. Viola and M. J. Jones, "Robust real-time face detection," *Intl. journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[18] L. Sweeney, "k-anonymity: A model for protecting privacy," *Intl. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[19] R. Gross, E. Airoldi, B. Malin, and L. Sweeney, "Integrating utility into face de-identification," in *Intl. Workshop on Privacy Enhancing Technologies*. Springer, 2005, pp. 227–242.

[20] R. Gross, L. Sweeney, F. De la Torre, and S. Baker, "Model-based face de-identification," in *Conf. on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*. IEEE, 2006, pp. 161–161.

[21] R. Gross, L. Sweeney, J. Cohn, F. de la Torre, and S. Baker, "Face de-identification," in *Protecting Privacy in Video Surveillance*. Springer, 2009, pp. 129–146.

[22] P. Korshunov and T. Ebrahimi, "Using face morphing to protect privacy," in *Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE Intl. Conf. on*. IEEE, 2013, pp. 208–213.

[23] ——, "Using warping for privacy protection in video surveillance," in *Digital Signal Processing (DSP), 2013 18th Intl. Conf. on*. IEEE, 2013, pp. 1–6.

[24] S. Tansuriyavong and S.-i. Hanaki, "Privacy protection by concealing persons in circumstantial video image," in *Proc. of the 2001 workshop on Perceptive user interfaces*. ACM, 2001, pp. 1–4.

[25] A. Williams, D. Xie, S. Ou, R. Grupen, A. Hanson, and E. Riseman, "Distributed smart cameras for aging in place," DTIC Document, Tech. Rep., 2006.

[26] S. B. Sadimon, M. S. Sunar, D. Mohamad, and H. Haron, "Computer generated caricature: A survey," in *Cyberworlds (CW), 2010 Intl. Conf. on*. IEEE, 2010, pp. 383–390.

[27] A. Hogue, S. Gill, and M. Jenkin, "Automated avatar creation for 3d games," in *Proc. of the 2007 Conf. on Future Play*. ACM, 2007, pp. 174–180.

[28] J. R. Padilla-López, A. A. Chaaraoui, and F. Flórez-Revuelta, "Visual privacy by context: a level-based visualisation scheme," in *Intl. Conf. on Ubiquitous Computing and Ambient Intelligence*. Springer, 2014, pp. 333–336.

[29] K. Chinomi, N. Nitta, Y. Ito, and N. Babaguchi, "Prisurv: privacy protected video surveillance system using adaptive visual abstraction," in *Intl. Conf. on Multimedia Modeling*. Springer, 2008, pp. 144–154.

[30] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. of the 27th annual Conf. on Computer Graphics and Interactive Techniques*. ACM Press/Addison-Wesley Publishing Co., 2000, pp. 417–424.

[31] H. Zhang and Q. Peng, "A survey on digital image inpainting," *Journal of image and graphics*, vol. 12, no. 1, pp. 1–10, 2007.

[32] A. R. Abraham, A. K. Prabhavathy, and J. D. Shree, "A survey on video inpainting," *Intl. Journal of Computer Applications*, vol. 56, no. 9, 2012.

[33] M. Granados, J. Tompkin, K. Kim, O. Grau, J. Kautz, and C. Theobalt, "How not to be seen - object removal from videos of crowded scenes," vol. 31, no. 2pt1, pp. 219–228, 2012.

[34] A. O. Akyüz and O. Kaya, "A proposed methodology for evaluating hdr false color maps," *ACM Trans. Appl. Percept.*, vol. 14, no. 1, pp. 2:1–2:18, Jul. 2016.

[35] S. Çiftçi, P. Korshunov, A. O. Akyüz, and T. Ebrahimi, "Using false colors to protect visual privacy of sensitive content," in *IS&T/SPIE Electronic Imaging*. Intl. Society for Optics and Photonics, 2015, pp. 93 941L–93 941L.

[36] ——, "Mediaeval 2015 drone protect task: Privacy protection in surveillance systems using false coloring," in *MediaEval Benchmarking Initiative for Multimedia Evaluation*, 2015.

[37] W. B. Pennebaker and J. L. Mitchell, *JPEG: Still image data compression standard*. Springer Science & Business Media, 1992.

[38] G. K. Wallace, "The jpeg still picture compression standard," *IEEE Trans. on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[39] J. Tesic, "Metadata practices for consumer photos," *IEEE MultiMedia*, vol. 12, no. 3, pp. 86–92, 2005.

[40] K. E. Spaulding, G. J. Woolfe, and R. L. Joshi, "Using a residual image to extend the color gamut and dynamic range of an srgb image," in *IS AND TS PICS Conf.* Society for Imaging Science & Technology, 2003, pp. 307–314.

[41] G. Ward and M. Simmons, "Subband encoding of high dynamic range imagery," in *Proc. of the 1st Symposium on Applied Perception in Graphics and Visualization*. ACM, 2004, pp. 83–90.

[42] ——, "JPEG-HDR: A backwards-compatible, high dynamic range extension to JPEG," in *ACM SIGGRAPH 2006 Courses*. ACM, 2006, p. 3.

[43] R. Mantiuk, A. Efremov, K. Myszkowski, and H.-P. Seidel, "Backward compatible high dynamic range mpeg video compression," *ACM TOG*, vol. 25, no. 3, pp. 713–723, 2006.

[44] A. Artusi, R. K. Mantiuk, T. Richter, P. Hanhart, P. Korshunov, M. Agostinelli, A. Ten, and T. Ebrahimi, "Overview and evaluation of the jpeg xt hdr image compression standard," *Journal of Real-Time Image Processing*, pp. 1–16, 2015.

[45] L. Yuan, P. Korshunov, and T. Ebrahimi, "Privacy-preserving photo sharing based on a secure jpeg," in *2015 IEEE Conf. on Computer Communications Workshops*. IEEE, 2015, pp. 185–190.

[46] M. Saini, P. K. Atrey, S. Mehrotra, S. Emmanuel, and M. Kankanhalli, "Privacy modeling for video data publication," in *2010 IEEE International Conference on Multimedia and Expo*, July 2010, pp. 60–65.

[47] M. K. Saini, P. K. Atrey, S. Mehrotra, and M. S. Kankanhalli, "Privacy aware publication of surveillance video," *International Journal of Trust Management in Computing and Communications*, vol. 1, no. 1, pp. 23–51, 2013.

[48] P. Deutsch and J.-L. Gailly, "Zlib compressed data format specification version 3.3," Tech. Rep., 1996.

[49] KitwarePublics, "Colormaps," http://www.itk.org/Wiki/Colormaps, [Online; accessed 21-May-2016].

[50] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[51] P. Korshunov, A. Melle, J.-L. Dugelay, and T. Ebrahimi, "Framework for objective evaluation of privacy filters," in *SPIE Optical Engr. Applications*. Intl. Society for Optics and Photonics, 2013, pp. 88 560T–88 560T.

[52] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. Computer Vision and Pattern Recognition*. IEEE, 1991, pp. 586–591.

[53] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *Patt. Analy. and Mach. Intel., IEEE Trans. on*, vol. 19, no. 7, pp. 711–720, 1997.

[54] T. Ahonen, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: Application to face recognition,” *Patt. Analy. and Mach. Intel., IEEE Trans. on*, vol. 28, no. 12, pp. 2037–2041, 2006.

[55] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, “The feret evaluation methodology for face-recognition algorithms,” *Patt. Analy. and Mach. Intel., IEEE Trans. on*, vol. 22, no. 10, pp. 1090–1104, 2000.

[56] B. E. Rogowitz and A. D. Kalvin, “The” which blair project”: a quick visual method for evaluating perceptual color maps,” in *Visualization, 2001. VIS’01. Proc.* IEEE, 2001, pp. 183–556.

Serdar Çiftçi obtained his B.Sc. degree from Department of Computer Engineering at Selcuk University in 2007. He received both his M.Sc. and Ph.D. degrees from the Department of Computer Engineering at Middle East Technical University (METU) respectively in 2011 and 2017. Currently, he is a teaching assistant at the Department of Computer Engineering at METU. Serdar is expected to work as an Assistant Professor at Harran University. His primary research interest is visual privacy protection.

Ahmet Oğuz Akyüz received his B.Sc. degree from the Department of Computer Engineering at METU in 2003. He obtained his Ph.D. in Computer Science from the University of Central Florida in 2007. He worked as a visiting researcher at Max Planck Institute for Biological Cybernetics in Tübingen, Germany in 2006. He is currently an Associate Professor at the Department of Computer Engineering at METU. His research interests are high dynamic range (HDR) imaging, color science, visual privacy protection, and rendering algorithms.

Touradj Ebrahimi is currently Professor at EPFL heading its Multimedia Signal Processing Group. He is also the Convener of JPEG standardization Committee. He was also adjunct Professor with the Center of Quantifiable Quality of Service at Norwegian University of Science and Technology (NTNU) from 2008 to 2012. His research interests include still, moving, and 3D image processing and coding, visual information security (rights protection, watermarking, authentication, data integrity, steganography), new media, and human computer interfaces (smart vision, brain computer interface). He is the author or the co-author of more than 200 research publications, and holds 14 patents.