Tag Detection for Preventing Unauthorized Face Image Processing

Alberto Escalada Jimenez[1], Adrian Dabrowski[2], Noburu Sonehara[3], Juan M Montero Martinez[1], and Isao Echizen[3]

[1] E.T.S. Ing. Telecomunicacin, Universidad Politcnica de Madrid, Madrid, Spain
alberto.escalada.jimenez@alumnos.upm.es
[2] SBA Research, University of Technology, Vienna, Austria
adrian.dabrowski@tuwien.ac.at
[3] National Institue of Informatics, Tokyo, Japan
iechizen@nii.ac.jp

Abstract. A new technology is being proposed as a solution to the problem of unintentional facial detection and recognition in pictures in which the individuals appearing want to express their privacy preferences, through the use of different tags. The existing methods for face de-identification were mostly ad hoc solutions that only provided an absolute binary solution in a privacy context such as pixelation, or a bar mask. As the number and users of social networks are increasing, our preferences regarding our privacy may become more complex, leaving these absolute binary solutions as something obsolete. The proposed technology overcomes this problem by embedding information in a tag which will be placed close to the face without being disruptive. Through a decoding method the tag will provide the preferences that will be applied to the images in further stages.

Keywords: Privacy, Face Detection, Tag Detection, Unauthorized Face Processing.

1 Introduction

Face detection and face recognition technologies have been highly developed during the last decades [1] [2] and they have achieved a high efficiency being capable of detecting faces at long distances or in low visibility conditions. As a representative fact, facial recognition systems are aimed to be installed in surveillance of public places, and access and border control at airports [3] while it is already a fact that for instance it is estimated that there are over 10,000 webcams focused on public spaces around the United States. This last fact was aggravated with the terrorist attacks of September 11, 2001, which immediately focused attention on the field of biometrics [4]. This development has been made only in favor of the technology and not taking into account its social implications, which means that the individuals whom their faces have being detected have not intervened in the process. This scenario has made that the individuals being recorded or

photographed have a few means to express their privacy preferences or their posture about being subjected to these new technologies. Lately, due to the improvement of portable devices with built-in cameras and the popularization of Social Networks in which photos are daily posted, like the case of Facebook, in which daily are uploaded 350 million photos on average [5], the widespread of a photography has a bigger impact than some years ago, and the access to these information is easier, and more out of the users control than what is generally conceived [6]. If we take a closer look at what some Social Networks use about face detection or recognition software, we find that Facebook , in which everyday people add more than 100 million tags to photos, offers a tag suggestion feature in which they use a face recognition software [7]. The facial recognition software uses an algorithm to calculate a unique number called template, based on some facial features appearing in the already tagged photos and profile pictures [8]. In this situation, a person who desires to block their faces in order to not being subject of a face recognition software has no means unless he is the one who has uploaded the picture.

In this paper we present a technology which allows the individuals that are being subject to a face processing method to express their privacy preferences through a visual code in which the information is embedded. In further steps, this flag-based system which is being worn by the individuals who want to protect their privacy under certain situations, is detected with a code detection algorithm, and thus the information contained can be extracted and applied to the image. These privacy preferences are set according to a policy framework presented by Adrian Dabrowski called Picture Privacy Policy Framework (P3F) which consists in a central database of privacy policies using a flag-based system [9]. In its work, the author developed a policy framework providing a solution to the problem of unauthorized face image processing. It covered all the aspects regarding how the desired system should work since the moment the policy was obtained from a code that the individuals would wear. That is the main difference with the work presented in this paper, which presents the a new novel technology that will be capable of encoding the privacy preferences of the user,and will therefore make this system feasible. An example of use of the P3F policy framework can be seen in Figure 1.

2 Previous Methods

In this section we will try to have a look at the two different approaches of methods that have been used for de-identification purposes: first the methods protecting privacy right before pictures are taken as a way of showing their attitude and preferences regarding their privacy in pictures, and then the methods that have been used for protecting the privacy of the individuals once a video or a photograph have been taken. We will also consider the commonly used and available visual encoding schemes and will explain in what they consist at the end of this section.

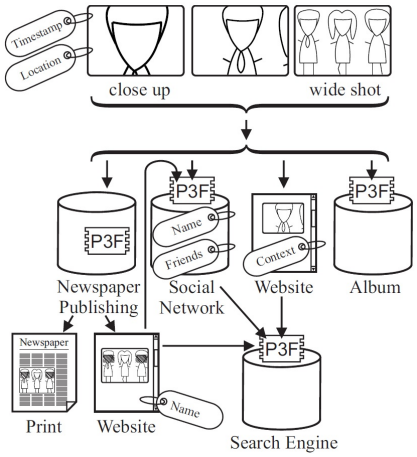

Fig. 1. Example where P3F sists in the distribution path [9].

As time passes, Social Networking Services (SNS) develop faster coming up with new ideas for sharing our lives and being connected to each other. That is the main reason why the preferences about our privacy become more and more complex and we need a new method as the one presented here to enable the individuals to express their preferences in different contexts and situations.

There are already some methods to protect the privacy and express the privacy preferences before the photos are taken. The most famous ones are called face mummification, but they are not successful because they are too radical and intrusive. The solutions for providing privacy once the pictures have being taken are called ad hoc and consist in making the face unrecognizable so the people appearing in a picture can maintain their anonymity. Some examples of ad hoc solution are blurring, pixilation, other distortion filters, or a more recent one called k-Same algorithm [10].

When we check the methods that already exist, we see that they consist in absolute binary solutions, in which the person decides, in a notably difficult way if it is possible (we always do not know where we have been photographed or recorded), whether they want their face in a picture or not, and this mechanism in today's more complex context as the SNS grow it is likely not to fulfill the users requirements and expectations. That is the reason why our solution outstands, it is an easy way to express more difficult decisions about privacy in an easy, non-time consuming, and less obtrusive way. As an example of the increasing complexity and the convenience of the hereby proposed technology, we can decide that we do not want to show our face to anyone with the same technology as if we were saying that we do not want to be recognized by any algorithm, but we still allow showing our face in certain SNS.

Some of the related work regarding already existing technology used for encoding information in a visual way are:

1. *1D Barcodes:* Linear bar codes are an obvious choice since it is a widespread technology and they are computationally easy to detect using frequency analysis. Nevertheless, the fact that they are a black and white technology might interfere with our face or eye detection algorithms. Since they are composed of several bars with different widths, their feasibility is relatively low when the distance to the code increases.
2. *2D Barcodes:* This is a more feasible scheme, since some of them can be customized almost from scratch (e.g. Microsoft Tag [11]). Most of them require an easily spottable wynchronization marker or a quite zone around them, and they are recognizable pattern which question the discretion required for such a technology, and thus questioning the validity of this method.
3. *Watermarking techniques:* Watermarking and stenography are often used in the context of digital rights management. These technologies are mainly used in digital surfaces and means, and therefore are difficult to implement in any physical device.
4. *Augmented Reality Markers:* This technology is similar to the 2D Barcodes, and furthermore it is highly resistant to distortion and enable calculation of the relative distance and angle of the barcode surface to the camera. On the other hand, they lack of the discretion, and probably, the unobtrusiveness requirements, since it is used and optimized for real-time applications.

There is a slight difference between the two kind of methods presented in first place which are technologies that have the purpose of protecting the privacy of individuals once their faces have been detected, and the methods presented later for embedding information in a visual platform, which are closely related to the technology presented in this paper.

3 Requirements of the Technology

Right before finding the perfect solution, we have to come up with a group of requirements that the desired technology has to fulfill. These requirements fall within 2 different categories, which will set a framework that may be used in future steps to compare different potential technologies as possible implementations for our system.

3.1 Technical Requirements

1. *Data Payload:* The code has to be able to embed a certain amount of information (6 bits if we are going to use the P3F scheme) which will express in a visual way the privacy preferences chosen by each individual appearing in an image. If the policies that are desired to be expressed increase, then the data payload has to increase as well.
2. *Redundancy:* This feature is desirable in our code since the more redundant the information in the code is ,the higher the chances of successfully detecting and decoding will be under worse conditions.

3. *Noise robustness:* The technology has to be resistant to the noise that can appear in a picture due to environmental causes such as smoke, or due to the camera device, especially in low-light and low-contrast situations.
4. *Filter resistant:* A lot of SNS offer a wide variety of popular filters that can be applied easily. For instance, Instagram (which offers the direct option of posting the image also in Facebook) offers 19 different filters which modify different aspects of the image appearance by adding a border or frame, making color space transformations to increase contrast, or make color cast [12].
5. *Blurriness:* The technology has to be immune to blurriness that can be added manually (again a popular feature offered by today's SNS), or because of external causes.
6. *Illumination:* Some pictures are taken in a low illumination environment such as night scenes and thus, the code is not exposed in the picture in desirable lighting conditions.
7. *Distance:* The code has to be detectable from long distances. This requirement comes from the big development that all face detection methods have been subject of. Now we have really accurate face detection methods and therefore, the code has to be also detectable and decodable for a successful technology.
8. *Compression stability:* Digital photography is subject to several compression algorithms which commonly destroy details in pictures. The most common compression method for photographs on the Internet is JPEG.
9. *Cropping Invariance:* Some user editing tools available in any smartphone device as well as some popular SNS offer the option of cropping pictures so it can fill a desired size or they just show a certain part of the picture. Therefore, our code has to be decodable even if it has suffered some cropping or some cut in it.
10. *Physical Support:* We need a place where we can put our code without being too annoying for the user. For a first approach in this technology, a frame of the user's glasses would be an ideal solution.
11. *Non Blind Decoding:* This requirement is desirable instead of a blind decoding because although the latter is more flexible, the former is easier to implement for a new research topic as it is the one presented in this article. The technology has to be identifiable with a database which contains all the possible options in which the code can take form (can be presented as).

All these requirements are only applicable if the face detection algorighm succeeds. That means that if for any reason the picture has not the appropriate characteristics so the faces are detected, our code does not have to be visible or fulfill the requirements either.

3.2 Aesthetic Requirements

1. *Unobtrusiveness:* When a subject decides to use the P3F technology, it has to be as simple as possible and does not require too much effort without causing any disturbance for the user.

2. *Location:* The location of the code has to be as close as possible to the face, since placing the code somewhere else would be useless when a picture just shows the face. We have also agreed that the code has to be placed always in the same location (that will be a requirement for the user), because that way the algorithm will be much more efficient and lighter.
3. *Discretion:* The technology has to be as discreet as possible since otherwise it will not be used. This might seem as a minor requirement, but it is not. For the feasibility of this technology, the tags have to cause the smallest discomfort to the user as possible, and also for other people's eyes.

In the proposed requirements we can see that there could be some requirements that seems opposite, for instance for fulfilling the distance requirement we could think of making a bigger code which would be seen from longer distances, but that would make the Discretion requirement to fail, since a big object close to the face seems to compromise the feasibility and success of such a technology. Therefore we cannot meet all of them although they are all desirable, and some of them will have more priority or importance than others.

4 Proposed Tag

For the tag that we are going to use to express our policy, we are going to use a novel way to put information through a visual way, and that is by using different sets of colors which will be scanned at the same time that a facial detection or recognition software takes place.

4.1 Color Space

The selected color space for the colors that will be used in the tag is the RGB color model, an additive color model in which red, green, and blue light are added in order to reproduce a certain color. There are many different ways of defining the different RGB colors through numeric representations, and we will use the digital 8-bit per channel representation, in which a RGB color is defined by three 8-bit numbers each of them corresponding to the red, green, and blue lights.

According to the P3F framework, we would need to embed 6 bits in our tag to fulfill the requirements. That is 64 different possibilities. We decide to use only 5 different colors of the RGB color space in order to use the strongest options maintaining a wide margin of distance in between the chosen colors in the RGB color space. These colors and their 8-bit representations are: Red (255,0,0), Green (0,255,0), Blue (0,0,255), White (255,,255,255), and Black (0,0,0). This way we are ensuring a robustness in our code through the use of five different and strong colors.

4.2 Tag Design

In order to embed 6-bit of data, which was the data payload needed to implement the P3F policy, using the five colors we already named we find a simple problem of combinations. With a 6-bit word, we can obtain 64 different words, and using 5 colors we need to know how many use in a combination of them to obtain a number equal or greater of possible combinations.

We have decided, for simplicity and creating a stronger code, that the order of the colors appearing in the tag will not be taken into account, and the information will only come from the colors appearing and not the place they occupy. Therefore, we come up with a simple problem of combinations in which we have to figure out how many colors to use from the set of 5 colors from the RGB color space so we can obtain a number equal or greater than 64 combinations. The equation used for combinations is (1) where r stands for the unknown number of colors that need to be used, and n is the set of 5 different colors selected. The result, is the number of possible combinations when r objects have been selected from a set of n objects.

Thus we obtain that placing 4 colors in a tag, we could obtain a number of 70 different combinations in which the order is not taken into account.

$$ {}_nC_r = \left(\frac{n!}{r!(n-1)!}\right) \tag{1} $$

We can see a first approach to this technology in the tag sketched in Figure 2.

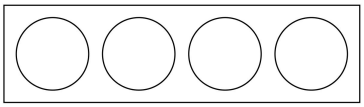

Fig. 2. Proposed tag which will be composed of a white background and four different circles in which will be placed the different RGB colors.

4.3 Tag Size and Other Features

In this section some of the physical features of the tag will be explained, as well as the criteria that is going to be used to base our decisions.

The size of the tag will be, in the beginning, such that all the aesthetic requirements are fulfilled, which means that it will have a width and length that will not be too intrusive in someone's face, but at the same time they will be big enough to be seen and processed correctly by the algorithm explained in this paper. We have already make a test plan, which includes testing how accurate is the processing of the tag according to its size and the size of the circles which appear in it (both sizes are closely linked). Therefore, we will be able to extract

an ideal size which will be transported to reality and thus, will optimize the results of this novel technology.

As it has already said, the colors used in this TAG belong to the RGB color space, and they are quite simple and robust since they are, or composed by, the extreme colors of the named space. This scenario makes the generation of the tags that will be worn extremely cheap and easy, since everyone with access to a computer can create a tag with the desired colors and with an optimal size which will be provided in the near future.

5 The Detection Algorithm

Once the technology has been implemented and the codes have been created , we will implement an algorithm which will be capable of detecting, locating, and analyzing the codes that appear in a picture. This algorithm is composed of two different parts that will lead to a lighter and more efficient search of the different codes shown in the picture. The phases and the explanation of the algorithm are given below, and an outlook of the whole algorithm can be seen in Figure 3.

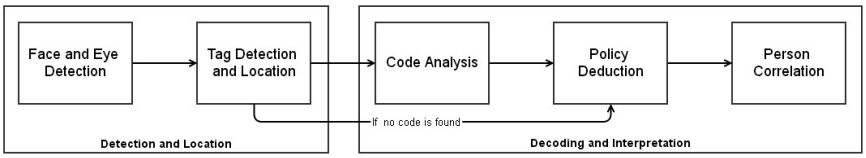

Fig. 3. Different stages of the detection and interpretation algorithm.

Furthermore, the algorithm proposal we are presenting here is composed of one more stage which does not depend on us. It is a watermarking phase which would be executed by the publishing entities, since they are responsible too of using the proposed algorithm for detecting and decoding the tags.

5.1 Detection and Location

In this project, for the computer image processing which will be hold during the whole development of the technology, it was decided to use an open source computer vision software such as OpenCV . This powerful tool will help us to analyze the different images subject of our research being capable of detecting and decoding the tags in which the information will be contained. We can differentiate two main stages in which this phase is divided.

The first step is to use a face detection algorithm in the image we are going to analyze, which will be critical to reduce the heaviness of the whole process. Since

the location of the code in the facial area (in a fixed place, as it was previously said) is an important requirement, we will first locate the faces appearing in a picture and this way we will reduce the area in which we will look for the code making the algorithm faster, lighter, more accurate, and more efficient. For this task, we have opted for the most commonly used method for face detection, which is the one reported by Viola and Jones in 2004 [13]. This method achieves highly accurate rates and high-speed detection since it is based on a multi-scale detection algorithm that uses cascade composition of the Haar-like features, image integration, and a cascade architecture with strong classifiers. By using a face detection method that uses Haar Cascades in OpenCV, we can obtain results similar as the shown in the second picture in Figure 4. Thus, being easier to reduce the area of searching to a smaller and lighter location close to one of the eyes of each face detected. If the image that is being analyzed does not meet some size requirements, the algorithm will resize it since we have proved that the eye detection stage has better results when the images are not so small.

As a result from the face and eye detection algorithm we obtain a matrix containing rectangles in which the faces are framed. The next step is the detection and location of the code. This step will be easy using the software already named. We first will use some geometrical approximations to reduce the area in which we are going to look for the code. Since the eyes and the face have been identified and placed, some parameters arise from that calculation, like the radius of the eyes, or face. Using those parameters, we create a rectangular space within the picture limits in which we frame both eyes, and we add a dynamic margin which will ensure that also the code is included in that rectangle. We use the radius of the face to create the margin, so if the face is bigger than expected, the margin will be bigger as well granting that the colors of the code will remain in the limits of the rectangle which will be subject of study. After that, a simple shape detection algorithm is enough for finding a rectangular shape. If this algorithm does not succeed, then we will just submit the whole rectangle to the proper color filtering process.

5.2 Decoding and Interpretation

Once the code has been placed within the limits of the image, the picture will be subject to some color filters that will determine which colors are appearing in our tag. If the tag has been detected in the shape detection part, then it will be easier that the colors are detected with a lower mistakes.

For detecting the different colors that appear in the tag, we will use again the OpenCV software that has been previously mentioned. Since the tag has a white background, we will just look for the other four RGB colors and unless four colors are detected in the tag, the white color will be supposed and there will be as many circles as the difference between the total number of circles, four, and the detected colored circles.

For detecting which circles are colored we will use a tool for detecting the existing colors between a certain range in the RGB color space, and since we have selected four different colors with wide margin we will use the mean of

those two extreme values to set the range limits. For instance, a colour will be considered black always that its three RGB components are greater than 128 in the 8-bit representation, that would be (128,128,128).

Once this step is complete the results will be compared with a data base in which all the policies are saved. We are using a non-blind decoding so all the possible codes have to be placed in the algorithm to finally obtain the information which is embedded in the tag.

5.3 Watermarking

With this schema we are trusting the publishing entities as the ones responsible of using this detecting and decoding technology in order to preserve the privacy of people. We can introduce a further step which will ensure that the privacy preferences information is maintained even out of the publishing entity space.

If we add a secure and invisible watermarking step to our schema, the privacy preferences now will be embedded in the picture and not just in the privacy code anymore. Therefore if any third party uses the picture and modifies it in such a way that the tag is not visible, or the code cannot be decoded properly, there will be still a secure way of knowing and proving what the individuals appearing in the picture opted for their privacy. This is step has to be provided by the publishing entity, in which we completely trust.

6 Evaluation

In this section we will present some of the results obtained in our attempt of using the already explained algorithm, and see how the results have been so far. In the following figures we will explain what are steps followed in the evaluation of a certain picture.

First we place a tag in a photo where a recognizable face appears.If we apply the face detection and eye detection methods, we obtain an image like the original one but with the eyes and the face framed in a circle, a consequence that they have been detected. once we have that, we can extract a smaller area as it is shown in the third picture in which we will look for our tag. All these three pictures can be seen in Figure 4.

In the following images in Figure 5, we can see the result of a number of transformations that the image suffer in order to determine the number of existing circles in the tag, and its colors.

Because we are adding the tag manually, and there is not time to analyze the picture under different and worse conditions, it is very likely that the accuracy will close to 100

After the paper is submitted, the research about this topic will continue and will have as an objective the quantification of its features. This means that the algorithm will be tested with different tag sizes, applying different filters to the pictures, changing the margins of colors, seeing how does all this work in a group picture, etc. And the results will quantify the quality of this proposed new technology.

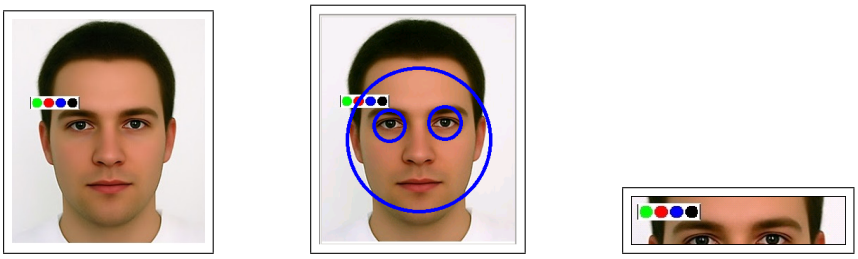

Fig. 4. The first picture shows the original photo with the policy tag. The second picture shows the photo after the eyes and face detection. And the third picture shows the tag area which will be scanned for extract information.

Fig. 5. In the image on the left we can see the contour of three out of the four circles. On the right we can see a method that detects green pixels.

7 Future Work

While the research was conducted, the potential of this technology was seen, but was not fully developed because of the time limitations. Because of this reason, we are proposing some guidelines that the research topic could follow in order to obtain a more successful technology.

A big future improvement that would improve the results of security preserved, and the immunity obtained in a picture would be to convert the whole system into a sensor-trusted one. That would mean that the decoding and application of each policy would be made right at the same time as the face detection takes place in any camera.

Another further step that could be taken in this research would be the implementation of the watermarking or steganography function that would provide, as it has already been explained, a more secure technology. That would be implemented also as a instant feature in the desired sensor-trusted improved technology.

Finally, a big improvement could be done for the success of this technology having to do with the appearance of the device that has to be worn. It seems that the discretion of this technology is a major concern for people, therefore, a new way of applying the already existing technology could be possible. For instance by placing different color spots within the limits of the frame of the eyeglasses, or using removable stickers that perfectly match the size of the frame. The only requirement here would be again the one asking for a detectable place.

8 Conclusion

The technology we have presented facilitates and allows the individuals appearing in any picture to express their preferences regarding privacy preferences within a framework already created.

The technique presented here makes the transmission of information through a novel way using colors located in a tag that will be worn in the proximity of the face. This technology enables a new way of expressing new and more complex preferences regarding the privacy of people appearing in pictures than existing previous methods.

A fast evaluation shows that the technology is feasible and that a possible quantification of its most important parameters would be ideal to improve it. Therefore, that is the direction the research will follow.

Acknowledgments. This work was performed under the National Institute of Informatics international internship program.

References

1. M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 1, pp. 34–58, 2002.
2. C. Zhang and Z. Zhang, "A survey of recent advances in face detection," Tech. rep., Microsoft Research, Tech. Rep., 2010.
3. A.-R. Sadeghi, T. Schneider, and I. Wehrenberg, "Efficient privacy-preserving face recognition," in *Information, Security and Cryptology–ICISC 2009*. Springer, 2010, pp. 229–244.
4. K. W. Bowyer, "Face recognition technology: security versus privacy," *Technology and Society Magazine, IEEE*, vol. 23, no. 1, pp. 9–19, 2004.
5. "A focus on efficiency," internet.org, Tech. Rep., 2013.
6. A. Besmer and H. Richter Lipford, "Moving beyond untagging: photo privacy in a tagged world," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010, pp. 1563–1572.
7. (2011, Jun.) Making photo tagging easier. [Online]. Available: https://www.facebook.com/notes/facebook/making-photo-tagging-easier/467145887130
8. (2014, May) How does facebook suggest tags? [Online]. Available: https://www.facebook.com/help/122175507864081
9. A. Dabrowski, E. Weippl, and I. Echizen, "Framework based on privacy policy hiding for preventing unauthorized face image processing," in *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, Oct 2013, pp. 455–461.
10. R. Gross, E. Airoldi, B. Malin, and L. Sweeney, "Integrating utility into face de-identification," in *Privacy Enhancing Technologies*. Springer, 2006, pp. 227–242.
11. M. Corporation. (2011) Microsoft tag - implementation guide. [Online]. Available: http://tag.microsoft.com/resources/implementation-guide.aspx
12. (2012, May) How does instagram develop their filters? [Online]. Available: http://www.quora.com/Instagram/How-does-Instagram-develop-their-filters
13. P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.