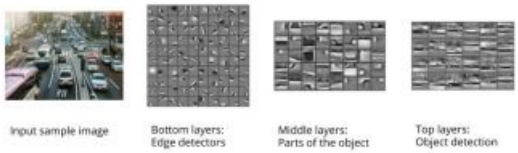
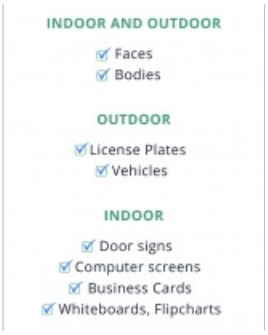
ACHIEVING LEGAL COMPLIANCE OF MOBILE MAPPING BY ANONYMIZING IMAGES

Image Anonymization for Mobile Mapping





Large amounts of personal data, such as faces and number plates, are collected during mobile mapping. Receiving consent from individuals for large-scale data collection is not feasible but is required by law in certain countries. Anonymizing images not only eliminates the need for consent but also increases trust in companies who conduct mobile mapping.

Privacy concerns in mobile mapping go back to 2010, when the Belgian Privacy Commission published the "Recommendation on Mobile Mapping".

The scope of the recommendation covered not only applications such as Google Street View but also other types of mobile mapping such as mapping by public authorities, mapping for tourism, real estate applications and GPS navigation mapping. In particular, it underlines that mobile mapping photos may involve the processing of personal data.

The inappropriate use of personal data has driven the public debate on data protection to an unprecedented level, leading to the GDPR regulation in May 2018. According to Art. 4 of the GDPR, personal data is defined as: "[...] any information relating to an identified or identifiable natural person". Considering that face and body parts are the most fundamental and highly visible elements of our identity, they fall under this definition, and consequently the GDPR. Similarly, number plates can be used to trace the identity of the subject, and so are also regarded as personal information.

Since GDPR is consent-based, the data can only be processed if the data subjects (i.e. persons) opt in. However, if you consider large imagery datasets with hundreds of thousands of individuals, this approach becomes time-consuming and costly.

An alternative solution comes with anonymization techniques. In fact, according to Recital 26, GDPR does not apply to anonymized data. Therefore, no consent is required for such data.

Current Image Anonymization Methods

Generally speaking, anonymization refers to any technique that irreversibly distorts an image in such a way that personal data cannot be reconstructed. Methods such as obfuscating a face with a swirl or some Photoshop effects may be able to deceive the human eye, but they are technically reversible; in other words the subject is re-identifiable, and therefore these methods are

ineffective.

Effective methods are:

- Replacing the affected area with a solid colour
- Pixelation, i.e. lowering the resolution of the affected area
- Blurring, which sets a pixel's colour value to the average of its neighbouring pixels.

For most mobile mapping use cases, blurring offers the best trade-off between performance, anonymization and reduced distortion.

The challenge is not the blurring itself, but detecting the affected objects. This is achieved using deep learning, through Convolutional Neural Networks (CNNs). These networks are capable of detecting structures and objects in images, such as faces.

Figure 1: Anonymization methods. Solid colour (left), pixelation (centre) and blurring (right). Image source: Unsplash.

Deep learning refers to the fact that the neural network is 'deep'; that is, it contains several layers of parameters and has to 'learn' the detection task by adjusting the parameters on some images where the type and position of the objects are known.

In CNNs, the layers close to the input image learn some basic features such as edges and corners. Based on these, the upper layers hierarchically extract more complicated features such as tyres or bonnets until the algorithm detects whole objects such as cars.

Unconventional Image Anonymization Methods

A novel technique is to apply another deep learning framework called GAN (Generative Adversarial Network) to replace objects like people or cars with artificially generated images. There are mainly two possibilities:

- Replacing objects with artificially generated objects
- 'Removing' the objects by replacing them with an artificially generated background

Both methods have their pitfalls. In the case of replacement with artificial faces, you might accidentally create one similar to a real person who could submit a data request, significantly increasing the administrative cost.

In the case of removing the objects, the generated imagery may be photorealistic but wrong. For instance, a car might cover some traffic signs or road markings in the original image, but in the generated fake background, these features are absent, thus damaging the validity of the whole dataset.

Additionally, GAN is more time-consuming and computing-intensive than blurring, casting doubt on its commercial viability for mobile mapping.

Figure 2: Layers of CNN. Image sources: Unsplash; Lee H. et al. (2011). Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks. Communications Of The ACM.

How Image Anonymization Should Be Done

Several aspects need careful consideration for the anonymization of mobile mapping imagery:

Quality of anonymization

Two important factors determine the quality: *false negative*, i.e. an object is not detected, and *false positive*, i.e. the background or an irrelevant object is wrongly marked as a relevant object. For anonymization, a false negative is in most cases more severe than a false positive: not detecting a clearly identifiable person poses a larger problem than, say, mistaking a construction container for a car.

However, depending on the use case, blurring a traffic sign because it is falsely identified as a number plate can be equally problematic.

Generally, good anonymization reduces both factors. Domain-specific know-how in mobile mapping in particular helps to improve the algorithm for street-level imagery.

What needs to be anonymized?

In principle, you should remove as much information as possible, but no more. The level of acceptable anonymization depends on the context. For mobile mapping, removing full bodies and cars is the safest option, but very often this is more than what many mobile mappers deem to be necessary. In most cases, only faces and number plates are required for anonymization.

Additionally, considering the possibility of false positives, anonymizing large areas increases the risk of removing important features. For instance, a statue might be misdetected as a person, in which case blurring only the face is a lesser problem than

blurring the whole object. Of course, with a top-shelf algorithm, you can reduce the risk of false positives.

Number of images and project deadlines

The time required for anonymization is linearly dependent on the number of images. For time-critical and large projects, cloud-based anonymization is often the preferred solution because of its high scalability, i.e. multiple computing nodes can work on anonymization in parallel.

Note that on-premise software cannot be scaled without purchasing and maintaining additional hardware and licences.

Price

Anonymization software that meets the quality standards for mobile mapping has high hardware requirements. The purchase and upkeep of dedicated hardware and the payroll cost for retaining staff with adequate machine learning know-how is prohibitive for most mobile mapping companies.

Cloud-based software offers a flexible and highly-scalable alternative.

Automation vs manual labour

In spite of the superiority of anonymization methods based on deep learning, some companies still outsource the task to contractors using manual labour.

Manual image blurring comes with certain problems and risks. It does not scale and can be the bottleneck for time-critical operations. Manual services are often provided in low-cost countries with weak data protection and labour laws where every person who comes into contact with the images increases the risk of data breaches.

Automated solutions significantly reduce these risks and problems.

Data protection and security due diligence

The anonymization provider of your choice must have measures in place to guarantee the safety and consistency of your data. If you or your customer is an EU-based company, they should have up-to-date documentation such as Records of Processing Activities and Technical and organizational measures (TOM) in place, as prescribed by the GDPR. There should be a dedicated Data Protection Officer (DPO) who conducts regular data protection audits. The processing should take place in GDPR-compliant data centres, where all storage devices with personal data should be encrypted. Furthermore, all subcontractors should be within the EU and comply with GDPR or have adequate Standard Contractual Clauses.

What needs to be anonymized?

Why Anonymization is Important for Society

Privacy is not only a regulatory requirement, but it is qualified as a **fundamental right** by the major international and regional human rights instruments, including:

- Article 12 of the United Nations Declaration of Human Rights (UDHR)
- Article 4 of the African Union Principles on Freedom of Expression
- Article 11 of the American Convention on Human Rights
- Articles 16 and 21 of the Arab Charter on Human Rights
- Article 21 of the ASEAN Human Rights Declaration
- Article 8 of the European Convention on Human Rights

Over 100 countries now have some form of data protection law. Nevertheless, it is still common that image technologies are used without particular regard to these protections.

Mobile mapping is one of the largest data collectors of imagery. For example, Google announced in May 2017 that they had captured more than 10 million miles (16 million kilometres) of Street View imagery across 83 countries. At the same time, mobile mapping could become one of the biggest sources of data misuse and breaches.

Conclusion

Mobile mapping is one of the rising stars in construction, surveying and geo-mapping because it is fast, accurate and easy to use. On the other hand, it has several privacy implications in terms of data protection and public opinion. Image blurring has been established as a standard anonymization method to overcome this hurdle. As mappers, we must consider whether anonymization complies with data privacy regulations and, as citizens, we must have the moral duty to protect individuals' identities.