Collecting and Processing Personal Data: Addressing Data Protection and Privacy Issues by Design

Marcos A Rodrigues and Mariza Kormann
GMPR-Geometric Modelling and Pattern Recognition Research Group
Communication and Computing Research Centre
Sheffield Hallam University
+44 (0)114 225 6911, +44 (0)114 225 6810
{*m.rodrigues, m.kormann*}@shu.ac.uk

ABSTRACT
The European Council has concluded in October 2013 that "*It is important to foster the trust of citizens and businesses in the digital economy. The timely adoption of a strong EU General Data Protection framework and the Cyber-security Directive is essential for the completion of the Digital Single Market by 2015*." This paper discusses the collection and processing of personal data in the context of the EU-funded ADMOS project whose aim is to detect in real time whether or not a person has noticed an advert and, if so, their profile in terms of gender and age. We review the issues and current legislation on the processing of personal data in the context of information systems, and how the project has incorporated a solution that fully satisfies current and proposed legislation on privacy.

Keywords
Personal data, privacy, image processing, gender classification, age estimation.

1. INTRODUCTION

The ADMOS Project (ADMOS, 2014) is funded by the EC and aims to develop a real-time gender classification and age estimation system to be used in private spaces for public use such as shopping malls, fairs and outdoor events. The purpose is to detect whether a person has noticed an advertising board and, if so, their profile in terms of age and gender. The *modus operandi* involves capturing live stream video and analysing each frame in real time from which statistical information are extracted concerning head count, gender and age profile. These statistics are then sent to a server at regular intervals as customised by the media owner (e.g. every minute, every hour, etc.). The main computing operations are thus the processing of live imagery performing face detection, gender classification and age estimation using appropriate algorithms (Rodrigues *et al.*, 2014a, 2014b). Note that the camera is placed on the advertising board itself, which means that if the person is looking at the board they are directly looking at the camera also.

The issue here is one of processing personal data as face images are classed as such. Specific regulations exist at European level concerning the right to privacy of individuals and these regulations are reviewed in the next section.

As an illustration, if we wish to use a camera pointing at the public for security purposes, special dispensations do exist as some of our individual rights and freedoms may be overridden by higher security concerns. In itself, this may constitute a dangerous set of circumstances to our civil rights, but it will not be discussed further in this paper as ADMOS is designed for marketing purposes. As such, the design must be fully compliant with our fundamental rights and freedoms encapsulated by relevant EU legislation.

First, it is important to disentangle the data. When a video frame is captured it contains personal data but as soon as it is processed the frame is discarded and the only information kept are the number of people who were facing the board in that frame, and the summary statistics concerning their age and gender. With this information it would be impossible to trace back to any living person. However, such simple solution would yield highly inaccurate and unreliable results, as the same person could be detected multiple times within a time frame of a few seconds. Consider the case in which a person might be looking at the board but stationary, or they might be looking at the board but moving fast or slow. To be of any use, the ADMOS design must be able to track a face from frame to frame but without performing face recognition, as face recognition would require a database (albeit temporary). Such approach would render the solution unworkable both on technical and privacy grounds.

This paper discusses the proposed solution within the framework of European legislation and has the potential to set standards for other projects along the same lines. In Section 2 we review privacy regulations within the European Union and its implications to digital products requiring the processing of personal data. In Section 3 we present the ADMOS solution and a discussion and conclusions are presented in Section 4.

2. PRIVACY REGULATIONS IN THE EUROPEAN UNION

It is convenient to highlight how a breach in privacy could cause harm to an individual in physical or financial terms. Hoven (2008) has identified four types of harm that may arise from misusing personal data as a result of privacy breach:

- Information based harm, with prime example being identity theft.
- Information inequality, where behavioural monitoring and analysis techniques may lead to personal discrimination on the basis of misleading or incorrect assumptions.
- Information injustice, where information presented in one context is used in another.
- Restriction of self-representation due to the omnipresence and pervasiveness of personal information, where multiple profiles exist across a number of different domains.

In addition there are also societal consequences that need to be taken into account. Indirect societal impacts include a sense of distrust or a loss in confidence or even fear in relation to those organisations using personal data.

The European Union legislative act created to regulate the protection of personal data is the Data Protection Directive of the European Parliament and of the Council of 24 October 1995, hereafter referred to as "the Directive". The Directive is a legal framework for the protection of individuals with regard to the processing of personal data and the free movement of such data. The legislation is part of the EU privacy and human rights law, defined in terms of fundamental rights and freedoms, notably the right to privacy of individuals. Member states are required to implement their own legislation but these cannot conflict with the European Directive.

The Directive's principles have set the standard for the legal definition of *privacy* and *personal data* and their regulatory uses in Europe and beyond. These include setting the scope of data protection, defining rights for data subjects and provisions concerning personal data and establishing supervisory authorities in the form of the EU level Article 29 Working Party (Robinson *et al*., 2009). The Art. 29 WP has an advisory status to the European Commission and acts independently. It is composed of a representative of the data protection supervisory authority (DPA) designated by each EU country, a representative of the European Data Protection Supervisor as a supervisory authority for EU Institutions and bodies (EDPS), and a representative of the EC.

The Directive principles are a unique legal instrument in how they support the exercise of a right to privacy and rules for personal data protection. Its principles include:

- Individuals should be informed when personal data are collected.
- Individuals should be told who is requesting the data and the reasons for their request to help them decide whether to release control of all or part of such data.
- Individuals should be told how access data about themselves in order to verify its accuracy and request changes.
- Individuals should be told how their data are protected from misuse.

In the context of information systems, the concept of *privacy* is directly related to how an individual may interact with and control access to information about themselves. Both the Universal Declaration of Human Rights by the UN in 1948 and the European Convention on Human Rights by the Council of Europe in 1950 recognise privacy as a fundamental human right. The concept of *personal data* and *controller* of personal data are clearly defined in Article 2 of the Directive. Personal data refers to any information that can be used to identify a person directly or indirectly; it could be an identification number, features specific to physical, physiological, mental, economic, cultural or social identity. A data controller refers to any natural person or legal entity which alone or jointly with others determines the purposes and means of the processing of personal data.

It is important to stress that the Directive is linked to the concept of personal data, and not to a notion of privacy. In this way, the Directive can be applied to cases of data processing that are not privacy sensitive on their own.

2.1 The Directive and the right to privacy of individuals

The provisions of the Directive's 34 Articles include data quality, special categories of processing, the rights of data subjects, confidentiality, security, liability and sanctions, codes of conduct and supervisory authorities. With reference to the particular issues raised by the ADMOS project, the main privacy goals and principles and how they are addressed by the Directive are highlighted in the following sub-sections under the headings of legitimacy, purpose and use restrictions, security and confidentiality, transparency, right to access and accountability.

2.1.1 On the legitimacy of data collection

Article 7 of the Directive defines criteria for legitimacy. Articles 7*a* and 7*f* are directly relevant to image capture and extraction of statistical information. They state that personal data may be processed only if the data subject has unambiguously given their consent, and if processing is necessary for the purposes of the legitimate interests of the controller or third parties, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject.

2.1.2 Purpose and use restrictions

Article 6 sets the purpose specification and proportionality use restrictions and also quality and accuracy requirements. It states that personal data must be processed fairly and lawfully, and it is the responsibility of the controller to ensure this. Data must only be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. Also personal data must be adequate, relevant and not excessive in relation to the purposes for which they are collected. Personal data must be accurate; when inaccuracies or incompleteness are identified data must be erased or rectified. Finally, personal data must be kept in a form that permits identification of data subjects for no longer than is necessary, and Member States can lay down appropriate safeguards for personal data stored for longer periods for historical, statistical or scientific use.

2.1.3 Security and confidentiality

Articles 16 and 17 set confidentiality and security safeguards principles for data processing. The confidentiality of processing states that any person who has access to personal data must not process them except on instructions of the controller, unless it is required to do so by law. Concerning security of processing, it states that it is the responsibility of the controller to implement appropriate technical and organizational measures to protect personal data against unauthorized access, disclosure, accidental damage or loss, or alteration. Any third party authorized to access and process personal data must do so only on the instructions of the controller governed by a contract or legal act binding the processor to the controller.

Note that personal data controllers are required to protect personal data against a variety of risks using appropriate measures. Such technical and organizational measures are normally incorporated at the design stage of data systems, so security becomes a built-in feature, a principle referred to as "privacy-by-design". This is a significant concept that must be incorporated into the ADMOS solution.

2.1.4 Transparency

The openness principle, the right to information regarding essential aspects of the data processing are set out in Articles 10 and 11. The controller must provide the data subject specific information including (even if the data have not been collected or obtained directly from the subject): the identity of the controller, the purpose of the processing, the recipients or categories of recipients of the data, information on the right to access and rectify the data. Some dispensation apply in case of processing for statistical purposes, historical or scientific research, or when providing such information proves impossible or would involve a disproportionate effort, or if it is lay down by law.

2.1.5 Data subject participation, right to access

The access principle is set in Article 12; access is normally coupled with the right to correct or delete personal data. The data subject has the right to obtain from the controller without constraint, excessive delay or expense: confirmation as to whether or not data related to them are being processed, communication in an intelligible form the source of the data, knowledge of the logic involved in any automatic processing of data. If data are inaccurate or incomplete, the controller must rectify the data. Finally, the controller must notify third parties to whom the data have been disclosed of any rectification, unless this proves impossible or involves a disproportionate effort.

2.1.6 Accountability

Articles 22 and 23 set out the accountability principle, or the rules on remedies and liability. Every person has the right to a judicial remedy for any breach of the rights guaranteed by the national law applicable to the processing of personal data. Any person who has suffered damage as a result of unlawful processing of personal data is entitled to receive compensation from the controller. The controller, however, may be exempt from this liability if they are not responsible for the events giving rise to the damage.

2.2 The new General Data Protection Regulation

New regulations have been proposed namely the General Data Protection Regulation (2012), whose aim is to harmonise the current data protection laws in place across the EU member states. The change from "Directive" to "Regulation" is significant, as it means it will be directly applicable to all EU member states without the need for national implementing legislation (as opposed to the Directive which required national legislation). Compared to the

Directive, the main changes under the new proposals are highlighted as follows.

- The "right to be forgotten" is reinforced.
- Explicit consent is required rather than assumed.
- People will have easier access to their own data.
- Increased responsibility and accountability of data controllers.
- People will refer to DPA-Data Protection Agency on their own country even when their data are processed by an organization outside the EU.
- EU rules will apply even if personal data are being processed abroad.
- A single set of rules on data protection will be valid across the EU.
- People will have to deal with a single national DPA in the EU country where they have their main base.

The reform is motivated towards stimulating growth in the digital economy especially for small and medium enterprises (SMEs). Having one set of rules instead of 28 of the Directive, it will cut costs and reduce red tape helping SMEs break into new markets. SMEs will be exempt from appointing a data protection officer as long as data processing is not their core business activity, it will remove the obligation to notify supervisory authorities, SMEs will be able to charge a fee for personal data access, and will be exempt from carrying out impact assessment unless there is a specific risk.

On 12 March 2014 the European Parliament approved the new Regulation with overwhelming majority. To become law, the proposed Regulation has to be adopted by the Council of Ministers using the *Ordinary Legislative Procedure,* which is the mechanism the vast majority of European laws are adopted jointly by the Parliament and the Council.

3. THE ADMOS SOLUTION

The above regulations were carefully taken into consideration in the design of the ADMOS client application. The main aspects are that all processing is done live in real time and no personal information is ever saved or leaves the system. An image of a face only exists in the application for the time duration it takes to process one single frame, after that all memory are erased. When statistical information is obtained on age and gender, this is saved for transmission to a server. Only the client can initiate transmission, and no external access is enabled.

3.1 Face detection, gender classification and age estimation

For each image, faces are detected through the Viola-Jones method (Viola and Jones, 2001, 2004) available from OpenCV libraries. An unconstrained image may contain a number of faces and each region of interest containing a face must be processed independently and the detected gender and age must be assigned to a corresponding data structure. The data structure thus, contains gender and age attributes such that to avoid unnecessary multiple calls to the gender and age classification functions if a particular tracked face has already

such definitions. This is clarified in the tracking of faces over multiple frames described in the next section.

Both gender classification and age estimation are based on LBP-Local Binary Patterns (Ahonen *et al*. 2006), (Pietikainen *et al*. 2011). LBP are grey-scale operators useful for texture classification defined over local neighbourhood pixels. It was originally defined using a kernel array of 3×3 pixels. The value of the centre pixel is compared with its neighbours and the result (greater or smaller) expressed as a binary number and summed over all pixels considered. LBP can be expressed over P sampling points on a circle of radius R where the intensity value of the centre pixel $I_{(x,y)}$ is compared to its neighbour pixels and summed over:

$$LBP_{P,R} = \sum_{p=0}^{P-1} T(I_p - I_c)2^p$$

Where I_p and I_c refer to the pixel intensity in the grey level of each neighbour and centre pixels respectively and T is a threshold function:

$$T(.) = \begin{cases} 1 & \text{if } (I_p - I_c) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Normally images are defined in blocks from which individual LBPs are calculated and then concatenated into a single histogram. The analysis of such histograms can be used to differentiate texture patterns. The size of the block under analysis can vary and this obviously will be reflected in the LBP histogram.

In order to classify whether an LBP histogram belongs to the class *Male* or *Female*, a simple binary decision rule is required. We use SVM-Support Vector Machines, which are algorithms that implement a mapping of pattern vectors to a higher dimensional feature space and find a 'best' separating hyperplane between the data set (Web and Copsey, 2011). Given a set of M training samples (l_i, x_i) where l_i is the associated class label ($l_i \in \{-1, 1\}$) of vector x_i where ($x_i \in R^N$), a SVM classifier finds the optimal hyperplane that maximises the margin between classes l_i:

$$f(x) = \sum_{i=1}^{M} l_i x_i \,.\, k(x, x_i) + b$$

Where: $k(x, x_i)$ is a kernel function, b is a bias and the sign of $f(x)$ is used to determine the class membership of vector x. For a two-class problem (*Male, Female*) a linear SVM would suffice. In this case, the kernel function is a dot product in the input space.

For age estimation, the problem here is to use SVM to classify an unknown person into one of the existing age groups. For example, group A: under 18, group B: 18—48, and group C: over 48. Three classifiers are defined for the different age groups using a 256-dimensional binary vector representing the LBP histograms. In order to reduce to a simple binary decision, three binary SVM classifiers are constructed: A versus (B,C), B versus (A,C), and C versus (A,B). Then assign a test pattern to the class (A or B or C) with the largest positive distance to the optimal separating hyperplane.

3.2 Data flow and multi-level queues

Figure 1 depicts the data flow for a single frame. The client application captures live image frame and detects all faces in the image. All faces that satisfy the condition of having one left and one right eye located in their expected regions of interest are considered to be looking at the advertising board and thus, are selected for further processing. Gender and age are estimated, and a tag is created for each face containing information such as current time and their exact location in the larger image. The tags allow tracking of faces over multiple frames anonymously. Tags move down a multi-level queue until they reach the last queue, when the processing of that tag is terminated, its information is saved for transmission to the server, and all memory are cleared.

Figure 1: *Flowchart for main threads for each grabbed image*

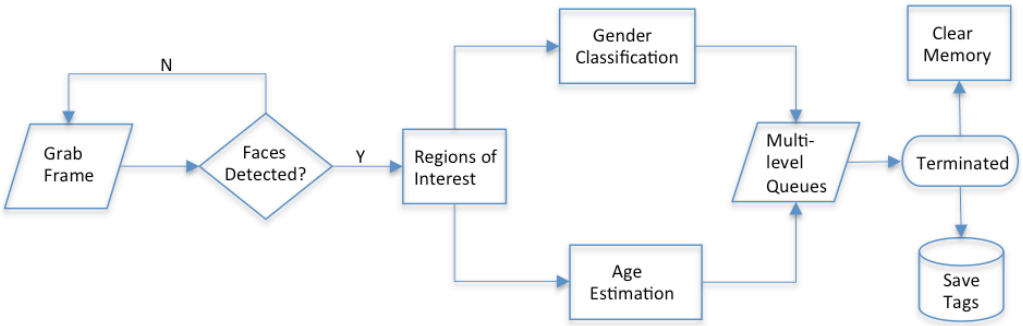
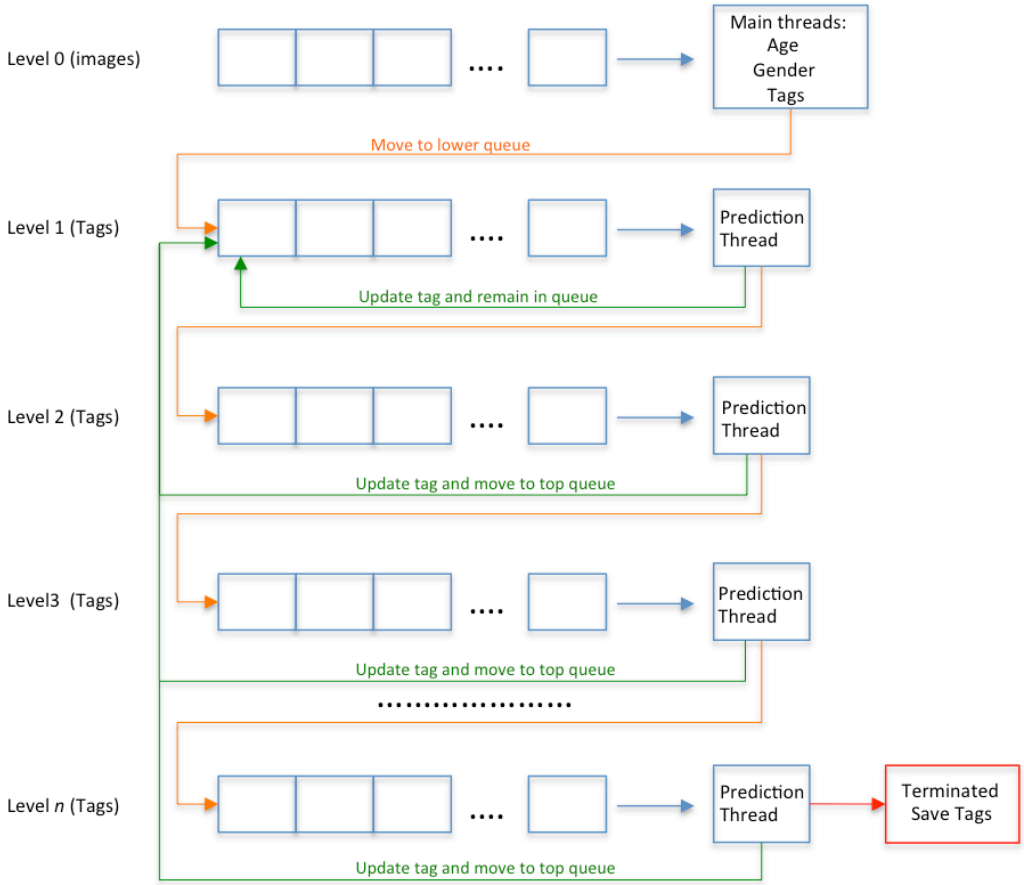


Figure 2 depicts the proposed thread based multi-level queue solution. The novel aspect of the solution is the design of a multi-level queue structure, which is controlled by multiple threads. This solution bears resemblance to MLFQ-Multi-Level Feedback Queue scheduling in operating systems (Mauro and McDougall, 2006). In multilevel scheduling, processes can move between queues as their priority changes. It can also prevent starvation by increasing the priority of processes that have been idle for a long time. When a priority is increased, a process moves one queue up, when it is decreased, one queue down. In our solution, the main difference is that a tag moves to a lower queue by default every time a new frame is processed. This condition can, however, be overridden to remain in the same queue (if already in the highest queue), or move to the highest queue (if it finds itself in any other lower queue). The consequence is that if a face tag is not detected in n consecutive frames, eventually it will reach the lowest queue and if it cannot be moved to the highest queue, then it is terminated. In other words, if the tag starves for n

consecutive frames it is terminated. Furthermore, in Unix there are 32 run queues, while in our design there is not such a limit. This is significant as the number of queues defines the number of frames over which a face can be robustly tracked and can be set relative to the speed of the processor.

Figure 2: *Multi-level queues for tracking anonymous faces*



Multi-level queue tracking is described as follows. At *Level 0* a tag is created for each face containing attributes for *Gender, Age, FrameNumber, StartTime, EndTime, StartLocation*, *EndLocation* and *Size*. A *Prediction Thread* runs through queues *Level 1* to *Level n* comparing the locations of existing tags with the new entrants. The closest distance between existing and entrants are estimated and, if less than a set maximum relative to its size then the new entrant is equivalent to the existing tag. In that case, the *EndTime* and *EndLocation* of the entrant replace the equivalent fields of the existing tag, which is raised to (or remains in) the top queue, and the entrant is fully discarded.

If an entrant tag has no equivalent tag in any level, then the (suspended) thread at *Level 0* will proceed to gender classification and age estimation placing the new tag in *Level 1*. This will save multiple calls to gender and age threads when this information has already been computed. Note that images are discarded at this point and only tag information is kept. Finally, all existing tags that had not their *EndTime* and *EndLocation* modified are lowered to the

next queue by default. When a tag reaches the last queue and cannot be raised to the top queue it is terminated and its information is saved for transmission to the server.

Figure 3: *Tracking anonymous faces over multiple frames*

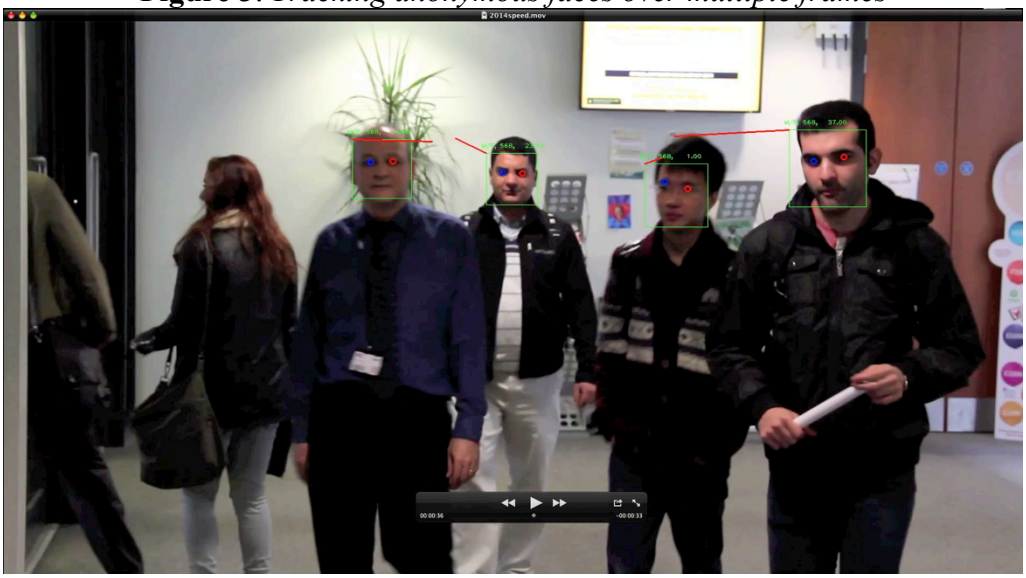


Figure 3 depicts the multi-level queue in action where tracking is annotated by a red line connecting the start location to the (current) end location of the top left corner of the face rectangle. All subjects are walking in direction to the camera; they started at a position near the back wall, which stands about 12m from the camera. The subject on the right has been successfully tracked continuously over 37 frames over a distance of about 8 metres.

4. DISCUSSION AND CONCLUSIONS

This paper has reviewed current and proposed European legislation concerning privacy and the processing of personal data in the context of the ADMOS project. Face images are considered personal data and require explicit consent. The consent issue is solved by restricting the use of ADMOS to private spaces for public use where specific public notices will be available with information on the identity of the controller, purpose of the processing, measures to safeguard personal data and so on.

Concerning the processing of personal data and protecting the fundamental right to privacy, the solution presented has significant properties: it involves real-time processing, faces are detected in an image but not recognised, only statistical information is retrieved, no images are ever kept in the system or transmitted to external devices, tracking over multiple frames does not require previous images, and any face image is only kept in memory for the time processing of a single frame. As soon as statistical information is obtained, all images are erased from memory. In this way, much of the personal data processing regulations do not apply for example security of the information held, legitimacy of data collection as no personal data are collected,

confidentiality, transparency, accountability, the right to access, request deletion or rectification of personal data, and so on. Moreover, from the statistical information contained in each tag it is impossible to trace back to any living person.

In conclusion, we have established sufficient safeguards with technical and organisational measures following a privacy-by-design principle that, together with the necessary information to be provided to the public are essential for the lawful deployment of the system. Therefore, the ADMOS solution fully complies with current European legislation and, as such, can be rightfully commercially exploited in the EU and beyond.

ACKNOWLEDGMENTS

This project has received funding from the European Union Seventh Framework Programme for research, technological development and demonstration under grant agreement number 315525, 2013–2015.

REFERENCES

ADMOS (2014): *Advertising Monitoring System Development for Outdoor Media Analytics*, EC Grant Agreement 31552 from 2013-2015. [Online] Available at http://admos.eu

Ahonen, T., Hadid, A. and M. Pietikainen M. (2006): Face description with local binary patterns: Application to face recognition. *TPAMI*, 28(12): 2037-2041

General Data Protection Regulation (2012): *Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 25.1.2012*. [Online] http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf

Hoven, J. (2008): Information Technology, Privacy and the Protection of Personal Data in Weckert, J., Hoven, J.; (eds) *Information Technology and Moral Philosophy*, Cambridge University Press 2008.

Mauro, J. and McDougall, R. (2006): *Solaris Internals* (2nd Edition), Prentice Hall PTR Upper Saddle River, USA, ISBN 0131482092

Pietikainen, M., Hadid, A., Zhao, G. and Ahonen T. (2011): *Computer Vision Using Local Binary Patterns*. Series Computational Imaging and Vision, Vol. 40, 212p, Springer

Rodrigues, M., Kormann, M. and Tomek, P. (2014a): ROI sensitive analysis for real time gender classification. In: Mastorakis, N, *et al*. (eds.) *Advances in information sciences and applications: Proceedings of 18th International Conference on Computers (part of CSCC'14)*. Recent advances in computer engineering series, 1 (22), 87-90

Rodrigues, M., Kormann, M., and Tomek, P. (2014b): A comparative analysis of binary patterns with discrete cosine transform for gender classification. In: Mastorakis, N, *et al*. (eds.) *Advances in information sciences and applications: Proceedings of 18th International Conference on Computers (part of CSCC'14)*. Recent advances in computer engineering series, 1 (22), 33-37

Robinson, N. Graux, H., Botterman, M. and Valeri L. (2009): *Review of the European Data Protection Directive*, http://www.ico.org.uk/upload/documents/library/data_protection/detailed_specialist_guides/review_of_eu_dp_directive.pdf

Viola, P., and Jones, M. (2001): Rapid object detection using a boosted cascade of simple features. In: *IEEE Conf. on Computer Vision and Pattern Recognition* (CVPR), pp. 511-518.

Viola, P., and Jones, M. (2004): Robust real-time face detection. *International Journal of Computer Vision* 57 (2), 137—154

Webb, A. and. Copsey, K. (2011): *Statistical Pattern Recognition*, 3rd edition, Wiley, 666pp.