20^{th} *Computer Vision Winter Workshop*
Paul Wohlhart, Vincent Lepetit (eds.)
Seggau, Austria, February 9-11, 2015

Biometry from surveillance cameras – forensics in practice

Borut Batagelj
Faculty of Computer and Information Science, University of Ljubljana
Večna pot 113, 1000 Ljubljana, Slovenia
borut.batagelj@fri.uni-lj.si

Franc Solina
Faculty of Computer and Information Science, University of Ljubljana
Večna pot 113, 1000 Ljubljana, Slovenia
franc.solina@fri.uni-lj.si

Abstract. *The article recounts various problems that the authors encountered in biometric face recognition and biometric image interpretation in their experience as court appointed expert witnesses. Before automated face recognition system can be applied on a typical surveillance video, images must be enhanced using various image processing methods or enriched by using computer vision 3D reconstruction methods. Authenticity of video material must also sometimes be verified. If face recognition is not possible or successful then other soft biometric characteristics can be checked. A legal expert witness for image biometry must be able to employ a large array of image processing and computer vision tools and methods. The expert witness must be able to explain how the biometric results were obtained, which were the necessary processing steps and how confident are the final results.*

1. Introduction

The multitude of image and video recording devices ranging from smart phones to the ever more extensive networks of video surveillance cameras produces a massive amount of imagery. Video surveillance is becoming ubiquitous in public and even private spaces. Therefore, the number of cases investigated by law enforcement, which have left some image related traces, is sharply increasing. When such cases subsequently enter some legal process, the need for expert witnesses with a working knowledge of image processing and computer vision is obvious. Interpreation of various security incidents recorded on video clips and photographs is beside the interpretation of material traces (fingerprints, bodily fluids etc.) gaining a steadily more central role in the judicial proceedings.

To correctly and independently interpret that imagery, expert witnesses are needed that can independently evaluate and interpret images. Often the main goal of such video interpretation is to identify or confirm the identity of a person. Researchers from our group have served now for several years as court appointed expert witnesses for interpretation of image and video material. In this article we would like to relate some useful experience from our practice. We discuss in the article only images where persons appear so that the tasks of the expert witness can therefore be consigned to biometry.

The tasks of an image biometry expert witness are much broader than just running a face recognition program [7]. Even if towards the end of the interpretation a face recognition system is used, several other actions on the image data must precede that step.

Since images are often recorded in suboptimal conditions, image enhancement methods must usually be applied (exposure adjustment, contrast improvement, noise filtering, stabilization of video etc.). The faces are often not captured frontally but from the side or from above so that standard frontal face recognition can not be applied directly. Building of a 3D face model is then usually attempted if images of the face from several views are available [3, 11]. Another problem can be a large age difference between the face images that we intend to compare and therefore some compensation for age related changes must be used. Another practical problem for comparison can be injury related abrasions or tattoos on the face of suspects. If face recognition can not be applied, then some other soft biometric features could sometimes be recovered. A person's height, for example, can be reconstructed even from a single image if enough other geometric information is available in the same image [5].

Sometimes the question of authenticity of the image material can arise. Has somebody tampered with the imagery to change the content of the material? There is a whole expertise area of image forgery detection, ranging from analysis of individual image elements, analysis of the image format, analysis of the input device and finally analysis of the physical and geometrical properties of the captured scene [6].

There are commercial software solutions that cover almost the entire range of forensic tasks [1]. However, the whole area of image biometry is moving so fast that a collection of different software tools, even open source tools, is often more flexible and usable. In any case, an image biometric expert witness must understand when and why certain processing steps or methods should be applied. A court appointed expert witness, in particular, must be able to understand and explain the whole process how he obtained and verified the results.

In this article we discuss only problems related to person identification using different biometric characteristics that we encountered during the past several years as court appointed expert witnesses. By discussing these cases we would like to illustrate the variety of biometric problems encountered in practice and the need to apply methods from a large range of different research results.

2. Face recognition

The most often posed question, that a court appointed expert witness is confronted with, is whether the accused person is really on the examined video clip? This is the problem of person verification. Usually, the expert witness has at his disposal a three-part mug shot from the police records and a video clip from a surveillance camera. Normally, only the face is used for identification. The courts expect that any face comparison should include a careful analysis of individual facial features and distances among them. We will discuss now the most common problems from practice.

2.1. Problems from practice

2.1.1 Poor image quality

Often the video quality of recordings from surveillance cameras is very poor due to low resolution and high compression rates. Such setting are usually chosen to save memory space on recording devices and only rarely due to the initial poor quality of the video signal itself. To save space, some surveillance systems are saving just a limited number of images per second or just images where some movement was detected. Due to all these circumstances, the quality of the video material is on numerous instances so poor that the application of advanced methods for face recognition that are based on facial features or on the integral face appearance is not possible [2].

2.1.2 Small scale face regions

A similar problem in digital face forensics, as poor image quality, is the insufficient size of the face region. The minimal interocular distance for reliable face recognition should be at least 32 pixels. Ideally, the interocular distance should be about 70 pixels. In practice, we often encounter images with a small resolution of 320×240 or 640×480 pixels where the face is furthermore often recorded from a larger distance. On such images the face region might have a size of only 15×15 pixels with interocular distance of mere 8 pixels. Even if the face is well illuminated and in frontal orientation, the success rate of face recognition systems is in such cases very low.

2.1.3 Non-frontal face orientation

Faces on surveillance video are often recorded from above and/or from the side so that the recorded face orientation is not frontal. Persons involved in criminal activity in addition try to evade the surveillance cameras and they tend to never look into the camera. All these circumstances add up to the fact that in the whole video recording of an event there is not even a single frontal image of a face. The faces of persons on such video footage are often partially concealed by sunglasses, hoods or caps which makes face recognition based on facial features even more demanding.

2.2. Possible solutions

Due to all the problems with image quality and face orientation described above, we try to use in such cases facial features that stand out even in images of poor quality. Such features are the shape of the head, the shape of the chin, the shape of the cheeks, the shape of the hairline and the baldness area, hair color, the shape of ears, nose and the size of the mouth. For recognition it can be beneficial also some irregularities or past injuries of the suspected person. Cases such as a nose deformation, a feature on the front of the adult men's neck (Adam's apple), excessive baldness or a prominent nose, all facilitate recognition. Facial or other visible tattoos can be very usable features for recognition even in images of very poor quality or resolution since they tend to stand out from the background of the skin color quite well.

The familiar feature on the front of the neck that is the forward protrusion of the thyroid cartilage.

2.2.1 Use of a face profile

When we try to analyze facial features on a video recording, it can turn out that the face profile is the most useful face orientation, because in the profile, certain features such as the shape of the nose stand out. As mentioned above, faces on surveillance video are captured from different often atypical viewpoints. This circumstance must be taken into account also in the case of profile views. If a suspect is available, the court can demand photos of the suspected person taken under different viewpoints, similar to those on the surveillance video.

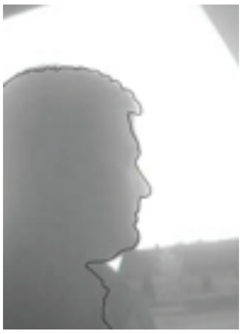

Figure 1. The silhouette of a person in front of an ATM.

The face profile is often usable in surveillance video from ATMs (Automated Teller Machines) where the face is normally backlit, making the face dark on a bright background. Although changing the exposure can help sometimes, often individual face features can not be made visible. Since a person in front of an ATM, who is performing an illegal activity, often looks around, his face profile is usually captured as a silhouette. Such silhouette can serve as a reference image for recognition from profile (Fig. 1).

2.2.2 Use of existing face recognition systems

Despite all the above described problems with different views and poor quality of video recordings, automated face recognition methods for frontal face recognition and from face sketches can be used. Before using such a method or a system, the input face image must be adjusted. Also, the results must be accordingly interpreted. To use a system for frontal face recognition, a 3D model of the corresponding face must be constructed from several viewpoints. The 3D model of the face is then used to generate the frontal view of that face which can subsequently be used as an input image for frontal face recognition [11].

A face recognition system can be used for comparison also on semi rotated faces, however, such a system must also be trained on similarly rotated faces. Another way of using existing systems is by drawing a face sketch or constructing a facial composite, based on the recorded video, and feed the resulting face to a face recognition system which can interpret also sketches [8].

Often offenders who are caught in the act are also suspects for other, similar, but unaccounted offenses. In such cases, the investigation needs to determine if two suspects are similar to each other. Systems for automatic face recognition are for such tasks also very useful.

3. Identification using other biometric features

Since face recognition is often not possible or not reliable enough, other personal features recorded in the surveillance video should be analyzed to help in the identification of a person. We will discuss the physical features of a person and his behavior. The following bodily features can greatly reduce the circle of suspects: body height [5], way of walking [9], way of handling objects and the actual body shape of a person. Most commonly, one tries to establish the body height of persons captured on surveillance video.

3.1. Estimation of body height

To determine the body height of a person on an image the Single View Metrology (SVM) can be used

[5]. Before applying the SVM method the image should be enhanced by increasing the contrast, improving the exposure and enhancing the edges. If several images of the same static scene captured under different illumination conditions are available the image can be improved by averaging the images similar as in high-dynamic-range imaging in order to sharpen and enhance the edges of the static objects on the scene. It is easier to derive the inherent geometric information (i.e. calibrate the space in all three dimensions (x, y, z)) from such enhanced images (Fig. 2).

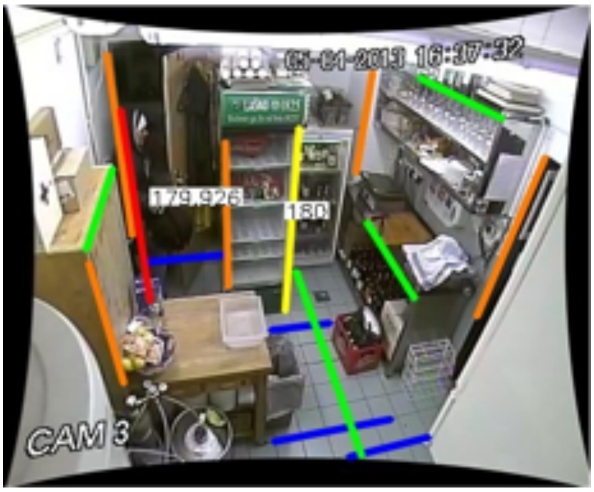

Figure 2. Calibration of a room should be performed after image distortions are corrected. The figure of the person in the corresponding video is never seen in its entirety. To be able to establish the height of the person, comparison to other calibrated lengths in the image is used.

Before calibration image distortions must be corrected so that objects on the image are correctly displayed. For calibration of the depicted space, portrayal of several rectangular objects aligned with the walls of the space is essential. Very useful for the calibration of the $x - y$ plane are for example quadratic plates in the floor paving. To determine the heights, the vertical axis z must be calibrated also. For this task one can use door and window frames or other vertical objects standing in the room (Fig. 2). Sometimes, if a room was rearranged in between, it is difficult to find a suitable reference object. Usually doors and windows are the most stable features of a room since they are seldom changed.

Sometimes, the video surveillance system was also changed in between. To perform a crime reconstruction or to determine the height of objects, images from the new system must be registered with the images from the old system, using objects that did not change in between. During the actual computation of the calibration one must enter actual measurement of known objects. Therefore a visit to the scene is necessary where as many objects as possible which are seen on the images should be measured to serve for the control of the accuracy of the calibration.

It is also very important that we use the original images when we do the calibration to be able to estimate the actual accuracy of the measurements. Accuracy depends on the resolution of the image and on the height of the person on the image. In normal circumstances, the error in determination of a person's height is about 5 cm. The SVM method therefore enables quite accurate determination of a person's height in an image. In special cases, when a person stands in the door frame or if we would like to estimate the size of an object such as a footprint, calibration of just two dimension of the space suffice, sometimes even just one dimension if the concerned object lies on a calibrated line.

3.1.1 Problems in a person's height determination

Determination of a person's height can be difficult if the person is not visible on the image in its entirety, for example, if the feet or the tip of the person's head are not visible. This can happen quite often if the camera is not mounted high enough or if the person is too close to the camera. In such cases, one can try to reconstruct the hidden body parts with the help of a general body model or the model of the observed person if the missing body part is seen in some other video frames.

Another often problem in determination of a person's height is that the person is on the entire video clip in a hunched posture due to running or brisk walking. If the person does not stop and straighten up, one must take this factor into account and determine at least the smallest possible height. For how much is the person taller, in addition to the determined minimal height, can be estimated using different phases of gait [10].

3.2. Other soft biometric features

If longer video surveillance clips are available the behavior of the observed person should be carefully analyzed. Walking has a certain personal character and can be used for identification [9]. Handling of

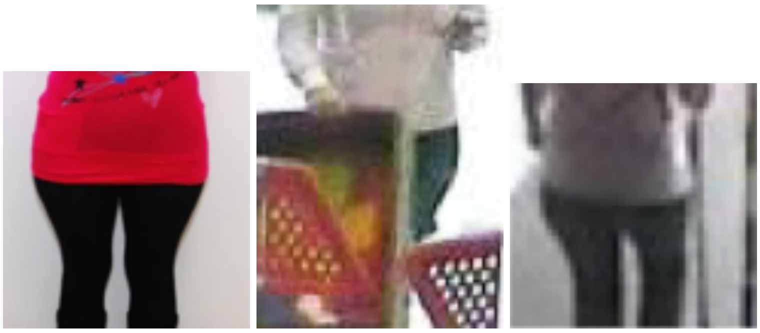

Figure 3. Wide hips can be used as a soft biometric feature.

objects can indicate the handedness of a person, for example with which hand one reaches products during shopping, with which hand one pays, reaches for cash on an ATM etc. It is also important how certain objects are carried, in which hand or over which shoulder one carries a bag. All those details can help in a person's identification. In cases of stealing of goods, one needs to check if it can be seen, that the suspect is hiding something under his clothing, or if his way of walking has changed. Any visible signs such as various inborn or injury related handicaps can also be used for identification. Tattoos are are well visible also on images of poor quality.

When we try to identify a person on a video clip, one should not concentrate only on the face features but also on the soft biometric features that can help us to reduce the number of suspects or to exclude a particular person from the list of suspects. Therefore, it is necessary to photograph for police records the entire body of a person where all particularities of that person can be seen. Fig. 3 (left) shows a person with disproportionally wide hips for the person's height. This size ratio can be verified on other images (Fig. 3, center and right).

If a suspect is apprehended immediately after a crime was committed, one can consider also features which can normally change quite rapidly, such as clothing, the shape and color of hair, existence of a mustache or a beard etc.

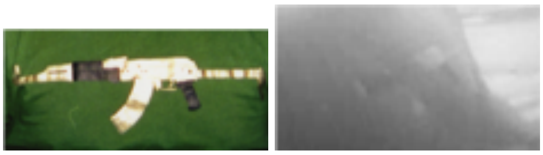

Figure 4. The design of a T-shirt (left) weared by a suspect, apprehended right after the crime took place, was identified on the surveillance video (right).

Clothing features can be used if the suspect could not have changed in between. Fig. 4 shows a case where the design of a T-shirt was used for identification. Special clothing features identified on surveillance video should be described in the report so that later, they can be searched for, for example, during a house search.

4. Other considerations from practice and experience

During a video surveillance system installation, one must mount the cameras so that the camera view angles cover the entire surveilled space and that the image quality is acceptable in all lightning conditions. All circumstances that might influence these two parameters should be considered. Sometimes, the surveillance system needs an extensive long time to adapt to sudden changes in illumination. The view angle of the camera can be obstructed by objects in the surveilled space. When some body parts of the surveilled person are occluded, the estimation of body height, for example, can be much harder. In the video corresponding to Fig. 2, for example, the figure of the person is never seen in its entirety, making the estimation of body height much more complicated.

Fig. 5 illustrates a poor placement of the surveillance camera, since when the door leading into the surveyed space is open, it occludes a large portion of the camera view angle, including the area where the vault was standing.

When the body height of a suspect is measured, it is very important to note if the person was wearing shoes or not. When analyzing events in front of ATMs, it is desirable if the clocks of the video surveillance system and the ATM system are synchronized. If not, then time intervals between ATM transactions should be used instead. Therefore, it is important to recover and save for analysis a much longer segment of the surveillance video where sev-

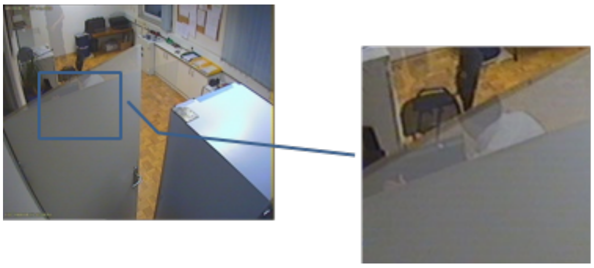

Figure 5. Bad placement of the surveillance camera–when the door is open, it occludes a large portion of the room, including the vault–the most important object in the room from a security viewpoint.

eral transactions are recorded. Then, several time intervals between transactions can be computed and based on the correspondence with the time intervals between ATM transactions, the timing in the video surveillance system can be aligned with the timing in the ATM system.

In some cases, it turned out, that video recordings from other nearby surveillance cameras would be useful, but it was already too late to obtain them. Namely, the legal obligation for safekeeping surveillance video is time limited, normally, at most up to three months, and then the old video data is usually erased by writing over new video data. Privacy advocates recommend the shortest legally required time for storing surveillance data and most producers of video surveillance equipment enable the storage of data between seven days and three months. Industry standards recommend that the storage capacity in a surveillance recording device should have a capacity to store at least 48 continuous hours of video with the recording parameters that enable a functional reconstruction of the events. For analysis of a crime event, public video surveillance footage can be also helpful, to determine, for example the escape direction or hiding of some material evidence.

Often the poor quality of video footage is a result of inappropriate copying of video data. The original video data can even get lost or stolen. In such cases, sometimes only images printed on paper remain. When original video digital data is not available and only poor quality printed images remain, advanced methods of image enhancement must be used [4].

5. Conclusions

Image material from video surveillance systems, which is used for face recognition, is often not suitable for direct use in automated face recognition systems. Images must usually be enhanced using a variety of image processing and computer vision methods. Sometimes even a manual step is necessary in the chain of recognition if software methods fail at a certain task. A professional sketch artist, for example, can draw a face based on video footage and the resulting sketch can be used as input into a face recognition system that is able to recognize also face sketches. If face recognition fails, then we can attempt to use other soft biometric properties, such as a person's height, for identification. An expert witness for face biometry must therefore have an understanding and a working experience of a very wide range of image processing and computer vision methods and tools.

References

[1] Amped software. http://ampedsoftware.com/five. online; accessed 19-November-2014. 2

[2] B. Batagelj and F. Solina. Face recognition in different subspaces: a comparative study. In *Pattern recognition in information systems : proceedings of the 6th International Workshop on Pattern Recognition in Information Systems, PRIS 2006 in conjunction with ICEIS 2006*, pages 71–80, Paphos, Cyprus, 2006. Insticc Press. 2

[3] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003. 1

[4] T. Bourlai, A. Ross, and A. K. Jain. Restoring degraded face images: A case study in matching faxed, printed, and scanned photos. *IEEE Transactions on Information Forensics and Security*, 6(2):371–384, 2011. 6

[5] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. *International Journal of Computer Vision*, 40(2):123–148, 2000. 2, 3, 4

[6] H. Farid. A survey of image forgery detection. *IEEE Signal Processing Magazine*, 2(26):16–25, 2009. 2

[7] A. K. Jain, B. Klare, and U. Park. Face matching and retrieval in forensics applications. *IEEE MultiMedia*, 19(1):2–10, 2012. 1

[8] B. F. Klare, Z. Li, and A. K. Jain. Matching forensic sketches to mug shot photos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):639–646, 2011. 3

[9] J. Kovač and P. Peer. Transformation based walking speed normalization for gait recognition. *Transactions on Internet and information systems*, 7(11):2690–2701, 2013. 3, 4

[10] J. Ljungberg and J. Sönnerstam. *Estimation of human height from surveillance camera footage-a*

reliability study. Examensarbete i ortopedteknik, Jönköping University, 2008. 4

[11] U. Park and A. K. Jain. 3D model-based face recognition in video. In *Advances in Biometrics, Proceedings 2nd International Conference on Biometrics, ICB*, pages 1085–1094. Springer, 2007. 1, 3