Android-Based Liveness Detection for Access Control in Smart Homes

Susanna Spinsante(✉), Laura Montanini, Veronica Bartolucci, Manola Ricciuti, and Ennio Gambi

Dipartimento di Ingegneria dell'Informazione, Universita' Politecnica delle Marche, 60131 Ancona, Italy
s.spinsante@univpm.it
http://www.tlc.dii.univpm.it

Abstract. In the domain of smart homes, technologies for personal safety and security play a prominent role. This paper presents a low-complexity Android application designed for mobile and embedded devices, that exploits the on-board camera to easily capture two images of the subject, and processes them to discriminate a true 3D and live face from a 2D one. The liveness detection based on such a discrimination provides anti-spoofing capabilities to secure access control based on face recognition. The results obtained are satisfactory even in different ambient light conditions, and further improvements are being developed to deal with low precision image acquisition.

Keywords: Liveness detection · Spoofing · Face recognition · Android · Stereo vision

1 Introduction

The Smart Home (SH) domain encompasses a huge variety of technologies, applications, and services, aimed at providing intelligence to an environment in which people spend most of their lifetime. Intelligent capabilities in SH aim at improving the quality of life of the resident people, by facilitating routine operations, and anticipating the users' needs, by learning and understanding their behaviours. Pervasive sensing plays a fundamental role in SH, as well as wireless technologies enabling the connection among heterogeneous devices, and creating the conditions for new integrated functionalities [1]. Among them, personal safety and security play a critical role [2], and many different devices and applications have been developed to address these needs. At the same time, however, regardless of how safe individual devices are or claim to be, new vulnerabilities may arise when different hardware devices are networked and set up to be controlled remotely.

Personal safety involves several aspects of the home security. Among them, the most important concerns the detection of alarming events, such as flooding,

O. Gaggi et al. (Eds.): GOODTECHS 2016, LNICST 195, pp. 116–124, 2017.
DOI: 10.1007/978-3-319-61949-1_13

gas and smoke leaks. Systems able to detect these events allow the user to be warned in time, promptly intervene and avoid potentially dangerous situations. Access control is another key point when dealing with security. Intrusion detection systems are becoming gradually more common. They exploit different technologies in order to detect the presence of unauthorized people at home. Most common technologies include magnetic sensors on doors and windows, motion sensors and cameras. Anyway, in this field, to distinguish between authorized and unauthorized subjects represents a primary objective.

Face recognition [3], one of the most successful image analysis and understanding applications, is a long-established research area that recently became extremely popular in consumer applications, thanks to advances in electronics and sensor technologies, that make high quality image sensors available in commercial devices, at a reasonable cost. Although extremely reliable methods for biometric personal identification exist, based on iris [4] or retinal scans [5], or fingerprint analysis [6], they still have to gain widespread acceptance by the general consumers.

In this paper, we address the design of a low-complexity Android-based application for liveness detection, based on image processing techniques, to be implemented in embedded Android platforms for video entry-phones. The aim of the project is to counteract face spoofing, one of the prominent threats to face recognition systems. The developed application focuses on the idea of discriminating a 2D face image, as a picture, from a real 3D face belonging to a live subject. The stereo vision technique is exploited, followed by the creation of a so-called *disparity map* in which, through different colors, the different areas of the image captured by the camera are highlighted, based on their relative distance from the camera itself. This way, the proposed application will discriminate a picture from a live face, according to the chromatic properties of the disparity graph output by the algorithm. A software library, named *BoofCV* [7], is used to implement the algorithm in a portable Android application.

The paper is organized as follows: Sect. 2 shortly presents the basic concepts upon which the application design has been conceived. The design of the Android mobile application is presented in Sect. 4, whereas Sect. 5 presents the experimental results obtained testing the application in real conditions. Finally, Sect. 6 concludes the paper.

2 Basic Concepts

For the human brain, binding an identity to a face is an automated and immediate task, despite its complexity. However, it is virtually impossible to reduce this operation to a search for objective parameters, that instead are essential to form the basis of an efficient biometric system. From a biometric point of view, we must consider that many factors can make it difficult to recognize a face, such as the different lighting conditions, the different facial expressions that a subject can take on, or the rotation of the face, the subject's age, physical radical changes, as well as the presence of obstructions, such as glasses, facial hair, or hair, covering part of the face. Despite these issues, algorithms that allow to obtain satisfactory results of personal identification have been proposed in the literature [8–10]. Nevertheless, some critical aspects have emerged recently, such as the resistance to external attack and, particularly, to spoofing. Anti-spoofing techniques applied to face recognition shall be unobtrusive, user-friendly, fast, low cost and well performing, e.g. able to avoid both false negative and false positive identifications.

Anti-spoofing, liveness detection and vitality detection are equivalent terms used in the literature to describe the same concept, that is: any technique aimed at verifying if the captured biometric information belongs to a live subject, or to an artificial and synthetic copy of him/her. As stated by Galbally et al. in [11], these techniques may be classified into three groups:

– sensor-level techniques: exploiting specific sensors in order to identify particular living traits (blood pressure, facial thermogram, etc.);
– feature-level techniques: in which the biometric data are acquired via a standard sensor and the distinction between fake and real faces is software-based;
– score-level techniques: much less common than the others, and focused on the study of biometric systems at a score-level.

The choice among one of the three groups should always balance advantages and disadvantages. Typically, hardware-based techniques have the best performance since they extract information directly from the human body. Nevertheless, they are quite intrusive and expensive. Conversely, the score-level techniques have limited performance, while maintaining low costs and intrusiveness. Among them, a compromise solution is represented by the feature-level group: it combines sufficient performance with low cost and less intrusiveness.

In the proposed application, the liveness detection problem is approximated as a problem of discriminating 2D from 3D objects, i.e. a picture of a face from a real face, exploiting a standard RGB camera. This is obtained basically by resorting to the stereo vision concept, according to which by comparing the images of the same subject captured from two different perspectives, the 3D information may be extracted, analysing the relative positions of the same elements in the two captured images. Since the application exploits a standard sensor and all the features are extracted by image processing, the technique used can be classified as a feature-level technique.

The robustness of the liveness detection process is then increased through a number of intermediate verification steps:

– check if the subject's nose is the element of the face at the shortest distance from the camera;
– check if different areas captured by the sensor (like nose and eyes, or face and background) are located at different distances from it;
– check if expected areas, like the face, the nose or the eyes, can be located in both the captured images;
– check if some kind of involuntary eye movements is detected.

3 Stereo Vision

The term stereo computer vision refers to the extraction of 3D information from digital images. By comparing the information captured from two different points of view, the 3D information can be extracted by examining the relative positions of objects in the two shots. Traditionally, in stereo vision, two different views of a scene are captured by horizontally disposed cameras: this mode is inspired by the binocular human visual system.

The problem of converting 2D information in 3D can be reduced substantially in two sub-problems: correspondence and reconstruction. The correspondence problem consists in identifying matched points in the images such that there are no ambiguities. In fact, ambiguous correspondences lead to different interpretations of the scene.

Figure 1 shows two simplified models of reality in order to better understand the concepts behind the stereo vision problem. In Fig. 1(a), two points P and Q on the same line of sight of the left image plane have been considered. Thanks to epipolar geometry the correspondence issue can be easily addressed. In fact, the epipolar constraint states that the projection of points P and Q in the right image plane must belong to the same line (dotted line) of their projection in the left side. The search in the space of corresponding points can then be narrowed from a 2D to a 1D search.

As regards the reconstruction, once the matching points have been identified, it is necessary to calculate their disparity. Referring to Fig. 1(b), the disparity can be defined as follows:

$$d = x_r - x_l. \tag{1}$$

It represents the difference between the x coordinate of the two corresponding points and allows to calculate the depth. In fact, through some simple steps, it is possible to obtain the relationship between the disparity d and the depth Z:

$$\begin{gathered} \frac{x_l}{f} = \frac{X}{Z} \;,\; \frac{x_r}{f} = \frac{X+b}{Z} \\ d = x_r - x_l = \frac{f(X+b)}{Z} - \frac{fX}{Z}, \end{gathered} \tag{2}$$

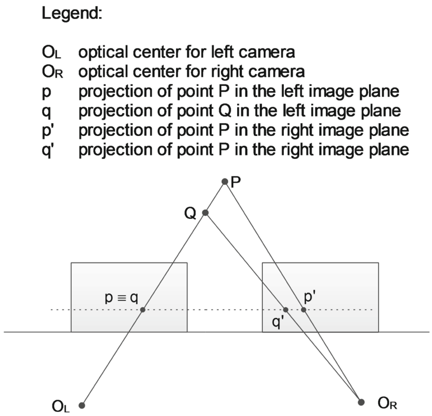
whence:

$$d = \frac{fb}{Z}. \tag{3}$$

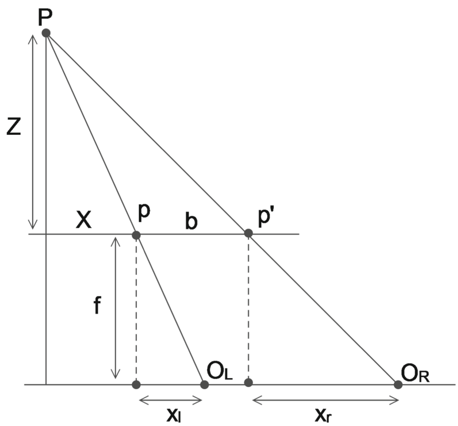
Therefore, the disparity of a point is proportional to focal length f and baseline b, and inversely proportional to its depth. Since f and b are constant over the whole image, a disparity map provides a direct encoding of the scene depth.

4 Design of the Android Mobile Application

A first version of the software application for liveness detection has been designed in Java language, for a desktop execution, exploiting the availability of the



(a) Epipolar constraint for correspondence: matches for p and q in the right plane must lie on the epipolar line (dotted line)



(b) A simplified model of reality in the top-view perspective

Fig. 1. Graphic representations of the stereo vision concepts

BoofCV libraries for stereo vision. Later, in order to get a portable code for Android devices, the *OpenBeans* library has been used. Using the camera sensor embedded in almost all the current mobile devices (smartphones), the application needs a couple of images gathered from two different perspectives (left and right), and processes them according to Sect. 2, to output a point-cloud of the detected subject. It denotes if a live (3D), or a fake (2D) picture, has been processed. The stereo vision implies a number of pre-requisites the captured images need to satisfy, such as: any image distortion due to the capturing sensor shall be compensated, in order to get pinhole camera - like images; each image in the couple shall be rectified, to be comparable. To this aim, two fundamental operations must be performed before starting the stereo matching and reconstruction processes: camera calibration and image rectification.

Calibration is a process for estimating the camera's intrinsic and extrinsic parameters. The first concern the internal characteristics of the camera, such as focal length or parameters of lenses distortion, while the extrinsic parameters describe the spatial position and orientation of the camera, i.e. the relative translations and rotations between the two images. The knowledge of the intrinsic parameters is an essential first step for the 3D reconstruction, because it enables the derivation of the scene structure in the space and removes the distortion of the lens, which leads to optical errors, degrading the accuracy. The *BoofCV* library provides a calibration feature. To calculate the parameters it is possible to use planar chessboards. Figure 2 shows a sample subset of the chessboard grids used to calibrate the device camera in our experiments.

Fig. 2. Sample subset of calibration chessboards

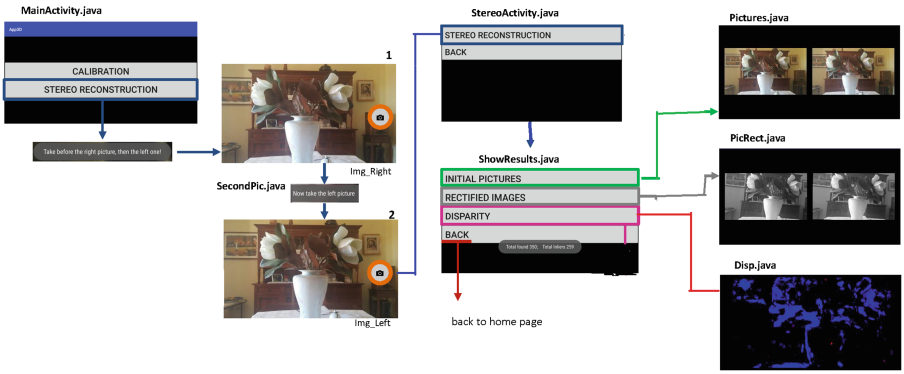


Fig. 3. Global flow diagram of the application steps

The image rectification is required when the considered image planes are not coplanar: thanks to this operation the images become coplanar, and the following procedures (matching and reconstruction) will be faster and more efficient, since they will run on a single dimension, as mentioned in Sect. 3.

In summary, three main steps are executed by the application: the camera calibration step; the rectification of the two images; the stereo matching between features belonging to the captured images, and, finally, the output point-cloud generation.

A global sketch of the different steps executed by the Android application is given in Fig. 3. The final image generated by the app features a single dominating color, which denotes a condition in which the input pictures are not associated to a live subject's face.

5 Experimental Results

In order to verify the proper functioning of the adopted method, several preliminary tests have been performed using a desktop PC. The objective of such tests is to identify the optimal resolution value which allows to correctly reconstruct the image. In fact, as the resolution increases, computational problems (processing time) increase or calibration inaccuracies appear. Test results show that the best resolution value is 0.3 megapixels. By using this value, the results are obtained in a very short time (about one second) and a correct calibration process is ensured. As already stated, this is a fundamental operation to properly carry out the stereo reconstruction.

Additional factors influencing the results are:

- the subject distance from the camera;
- the relative distance between the two pictures;
- the brightness.

For each of them, three different conditions have been considered, as shown in Table 1, and every possible combination of them has been considered. Results suggest that using as input pictures taken at a great distance produces worst results, at any brightness condition. For minimum and medium distances, the low-light condition does not provide acceptable results in any circumstance, while in the normal light case the face is entirely detected and reconstructed, both for extremely close-up, close-up and half-length photos. For the intense brightness case, acceptable results are obtained only if the two shots are very close together. In summary, in order to obtain a proper 3D reconstruction of the face:

- it should be well lit;
- the use of flash should be avoided, unless the two shots are taken at a distance smaller than 5 mm;
- the pair of photos should have a relative distance varying from 3–5 mm to 15 mm;
- within the limits of the considered situations, the distance of the subject from the camera is indifferent for a correct result.

After preliminary tests, the Android application has been tested on different devices (Samsung Galaxy S3, S5 and S6, LG G4, Huawei P8 Lite), equipped with diverse Android OS versions. The app execution runs smooth, if the device has at least 1 GB RAM available, but in any case it is necessary to downscale the camera resolution by a factor of at least 8, to enable a real-time processing of the captured images.

Also in this case, despite the not full precision of the acquisition process, the mobile application is able to discriminate faces shown in pictures (spoof) from live ones. Figure 4 summarizes some of the results obtained from lab experiments.

Table 1. Summary of the situations envisaged in the test phase, for each analysed factor.

Factor	Possible conditions	Description
Distance between subject and camera	Extreme close-up photos	The pictures contain just part of the face and neck
	Close-up photos	The pictures contain not only face and neck, but also the shoulders
	Half-length photos	The picture is cut at chest level
Distance between the two pictures	Minimum distance	The photos are almost coincident and the distance between the two shots is about 3–5 mm
	Medium distance	There is a slight shift between the two pictures (∼15 mm)
	Great distance	The two pictures are far apart more than 30 mm
Brightness	Minimum brightness	The subject is illuminated by a low light (for example an abatjour)
	Medium brightness	The subject is illuminated by a common warm light bulb
	Intense brightness	The subject is illuminated by a common, warm light bulb and in addiction by a flash

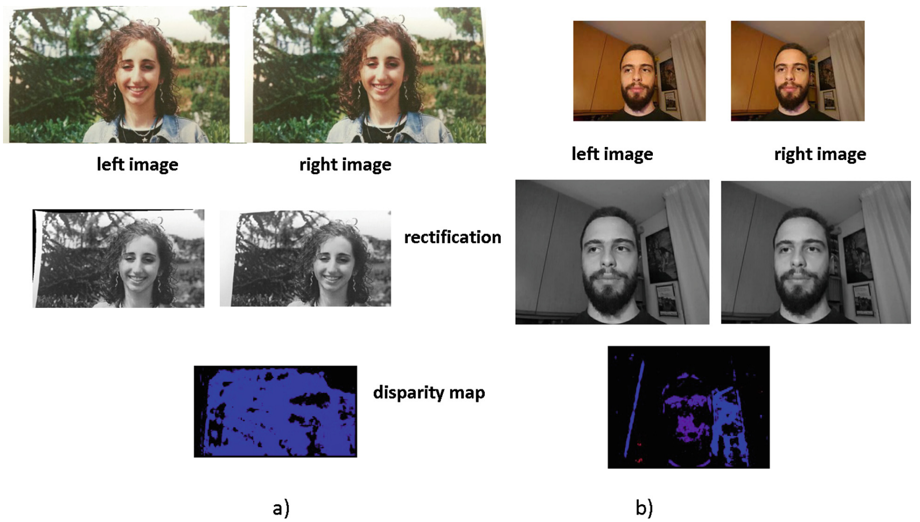


Fig. 4. Output of the liveness detection mobile app in the case of: (a) spoofed face images, (b) live and complex face image. The output disparity map in (b) clearly features the face profile distinguishable from the background.

6 Conclusion

The Android application for mobile and embedded devices presented in this paper demonstrates the feasibility of a real-time liveness detection process, which implements anti-spoofing by detecting the 2D or 3D nature of the captured face images. The application requires initial calibration of the device camera, and suffers from limitations due to possible imperfections in the image capturing process, however, it shows very promising results. Further developments are currently ongoing to increase the robustness of the application and test it on a larger variety of real-life conditions.

References

1. Spinsante, S., Cippitelli, E., Santis, A., Gambi, E., Gasparrini, S., Montanini, L., Raffaeli, L.: Multimodal interaction in a elderly-friendly smart home: a case study. In: Agüero, R., Zinner, T., Goleva, R., Timm-Giel, A., Tran-Gia, P. (eds.) MONAMI 2014. LNICSSITE, vol. 141, pp. 373–386. Springer, Cham (2015). doi:10.1007/978-3-319-16292-8_27
2. Ur, B., Jung, J., Schechter, S.: The current state of access control for smart devices in homes. In: Proceedings of Workshop on Home Usable Privacy and Security (HUPS) 24–26 July 2013, Newcastle, UK, pp. 1–6 (2013)
3. Zhao, W.Y., Chellappa, R.: Image-Based Face Recognition: Issues and Methods. Image Recognition and Classification, pp. 375–402 (2002). (Edited by B. Javidi, M. Dekker)
4. Gragnaniello, D., Sansone, C., Verdoliva, L.: Iris liveness detection for mobile devices based on local descriptors. Pattern Recogn. Lett. **57**(1), 81–87 (2015)
5. Azemin, M.Z.C., Kumar, D.K., Sugavaneswaran, L., Krishnan, S.: Supervised retinal biometrics in different lighting conditions. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, pp. 3971–3974 (2011)
6. Callaly, F., Cucu, C., Cucos, A., Leyden, M., Corcoran, P.: Real-time fingerprint analysis & authentication for embedded appliances. In: 2007 Digest of Technical Papers International Conference on Consumer Electronics, Las Vegas, NV, pp. 1–2 (2007)
7. BoofCV Software Library. http://boofcv.org/index.php?title=Download
8. Wang, J., Lu, C., Wang, M., Li, P., Yan, S., Hu, X.: Robust face recognition via adaptive sparse representation. IEEE Trans. Cybern. **44**(12), 2368–2378 (2014)
9. Wagner, A., Wright, J., Ganesh, A., Zhou, Z., Mobahi, H., Ma, Y.: Toward a practical face recognition system: robust alignment and illumination by sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. **34**(2), 372–386 (2012)
10. Liao, S., Jain, A.K., Li, S.Z.: Partial face recognition: alignment-free approach. IEEE Trans. Pattern Anal. Mach. Intell. **35**(5), 1193–1205 (2013)
11. Galbally, J., Marcel, S., Fierrez, J.: Biometric antispoofing methods: a survey in face recognition. IEEE Access **2**, 1530–1552 (2014)