2018 International Conference on Cyberworlds

Kinect vs Lytro in RGB-D face recognition

Valeria Chiesa
EURECOM
450 Route des Chappes,
06410 Biot, France
Email: valeria.chiesa@eurecom.fr

Jean-Luc Dugelay
EURECOM
450 Route des Chappes,
06410 Biot, France
Email: jean-luc.dugelay@eurecom.fr

***Abstract*—Light field cameras are becoming increasingly popular thanks to higher capabilities with respect to regular cameras in capturing information of a scene. Even though the principle associated with structured light sensors is quite different from the technology behind light field cameras, data provided by these technologies are similar in terms of depth map. With the aim of comparing the potential of Kinect and Lytro sensors on face recognition, two experiments are conducted on separate but publically available datasets and validated on a database acquired simultaneously with Lytro Illum camera and Kinect V1 sensor. The results obtained on RGB and depth maps are integrated with an experiment based on fusion at score level. The introduction of depth information in the RGB data is found more effective than standard bi dimensional imaging, especially in case of occlusions.**

***Keywords*-Face recognition; RGB-D images; Lytro; Kinect**

I. Introduction

Automatic recognition of a person by physical characteristics has become of primary interest. Biometric systems are now adopted at border controls, security accesses, and customized advertisements. In the last years, with the aim of improving the performance of biometric system, several researches investigated the influence of different acquisition technologies [1], [2].

Particular interest has been paid to structured light technology implemented in Kinect V1 cameras for low cost-effective multimodal sensing. Kinect [3] is a camera that combines normal images with IR reflection information, providing RGB-D data to the end user. In [4], a database of 1581 images from Kinect sensor has been created with the purpose of tackling the problem of face detection. In [5], authors presented an algorithm that uses 3D sensor for face recognition under challenging conditions and test its performance on a Kinect database. Information stored in depth map is proven useful to improve face recognition [1], [6] and novel efficient LBP-based descriptors have been designed [7]. Several experiments on anti-spoofing with Kinect images have been also done: in [8] a 3D Mask Attack Database is introduced and processed, showing how LBP features are powerful tools to detect 3D spoofing attacks.

At the end of the first decade of 2000, light field (or plenoptic) cameras have emerged on the market thanks to companies like Lytro[1] and Raytrix[2]. The number of publicly available databases is yet to become important [9]. Only a few studies on biometrics using light field cameras have been performed so far. Some of them are based on light field images property of being refocused as post-process: different focus levels can actually be systematically exploited as described in [2] where authors generate a fusion algorithm obtained by combining images rendered at different focal distances from the same light field data for face recognition. Others explore the usefulness of this technology in relation to Presentation Attacks Detection (PAD) [10], [11]. The third dimension information is added to normal pictures and in [10] a preliminary analysis for comparison with normal pictures is presented.

The fusion of biometric measures collected from the same person could improve the recognition rate respect with the single acquisition. The integration of complementary information has been extensively studied [12], [13]. In particular, fusion at score level is often preferred over other methods in multimodal biometric systems [14], [15]. It is less computationally expensive and more versatile than other kind of fusions.

The improvements due to the integration of RGB and D information in face recognition has been already proved in literature [16]. However, a straightforward comparison of the two technologies on the same biometric problem has not yet been launched.

The goal of this work is to carry out a systematic study of comparison of Lytro and Kinect camera in face recognition. This could give insight whether or not the differences between the technologies play a significant role in adopting one device or the other in the problems faced in biometrics.

In the first section databases used are presented. The second part is dedicated to pre-processing, baseline techniques and configuration description of the experiments. Then, the results of the tests are commented and integrated with a consistency analysis. Finally, a preliminary study of depth map quality is described and conclusions are discussed.

[1]Lytro, Inc. ”www.lytro.com”
[2]Raytrix, Gmbh. ”www.raytrix.de”

II. DATABASES

In order to compare the performances of face recognition algorithms on Lytro and Kinect V1 images, two publically available databases are used: IST-EURECOM Light Field Face Database (LFFD) [17] and EURECOM Kinect Face Database (KFD) [1]. The first is chosen because, as far as we know, it is the only light field database suitable for face analysis that includes depth information [18]. Although if bigger databases exist [4], KDF is selected because of its similarity with the first one. Both are acquired in a controlled environment, both consist of two sessions and have several facial expressions and examples of occlusion in common. A database composed of 20 subjects is acquired with both Lytro Illum camera and Kinect V1 sensor at the same time. The purpose of this home made database is to prove that the conclusions attached to this work are not influenced by environment or subjects variation in databases. In fact, all considered databases have been collected in similar environment. We chose this particular kind of data in order to minimized as much as possible the impact of external factors like not uniform illumination or particular background. The relatively small size of KFD database has not prevented in the past from carrying out similar analysis [1].

III. EXPERIMENTS

A. Pre-processing

Images from JMD, KFD and LFFD have been pre-processed in order to enable the analysis described in the next section. Both RGB and D images are aligned and cropped using the Python Package DLIB [19].

B. Techniques and configuration

With the aim of studying the potential of light field and structured light camera, three standard methods designed for RGB images (PCA [20], LBP [21] and LGBP [22]) and one customized for depth images (LBP3D [23]) are tested.

Two experiments are conducted on LFFD and KFD separately. For both experiments and both databases, six significant face variations are selected: neutral face (only for the second session), one emotion (smiling face), one action (open mouth), one different illumination (high illumination level) and two occlusions (hand on mouth and sunglasses). The experiments are designed to study the distance between test samples and gallery samples. In experiment E1, only the neutral faces acquired during the first session is considered as enrollment set, while in experiment E2, neutral expression for both the sessions is used as gallery. For each image in the test set, the closed sample from the gallery is associated and the ratio of correct matching is used as evaluation. Finally, fusion between RGB and D images is computed at score level for each method and results are compared with baseline generated.The same experiments have been carried out on RGB images with different resolutions and on smoothed depth maps. Since the results do not significantly move away from the described performances, they are not reported.

C. RGB images

RGB images from both LFFD and KFD database are analyzed (tab I and tab II).

Experiment E1: the ratio of good matching evaluated on pictures acquired during the first session is higher than the others for both LFFD and KFD. The gallery composition could explain this phenomenon: in experiment E1, enrollment set is composed by images illustrating ”Neutral” expression recorded during the first session. In addition to that, the time gap between the two sessions in KFD and LFFD databases is at maximum 14 days and at least 1 month respectively. This difference gives rise to imbalanced results between databases. In the case of ”Sunglasses” occlusion, the phenomenon is particularly evident: the recognition rate on LFFD is 51% using LBP-based method, but in KFD it is higher than 88%. For both KFD and LFFD, PCA-based method does not suit in case of occlusions produced by sunglasses or hand on face, while, using LGBP-based method, recognition rate is always higher than 70%.

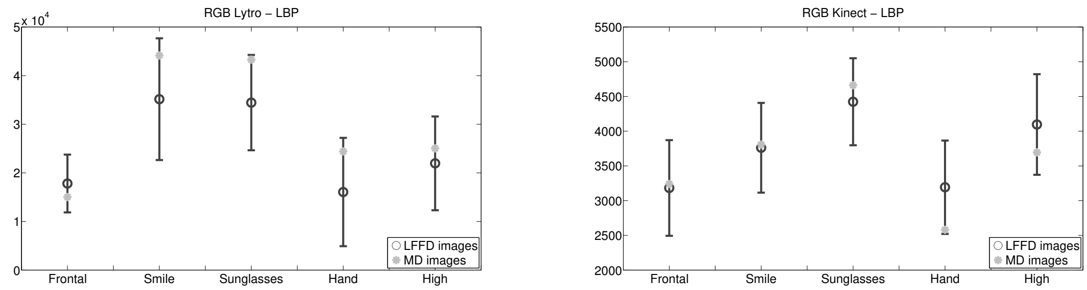
Experiment E2: here, the enrollment set is composed of two images for each subject (one for each session). Tests done on images of the first session reveal better performances with respect to experiment E1 for both LFFD and KFD. The time gap between acquisition sessions influences no longer the analysis. Considerations done for experiment E1 about PCA-based method are still valid: eigenface features are not tailored for recognition in case of occlusions.

D. Depth images

1) Depth images: preliminary analyses: Because of the different technologies used, Lytro and Kinect V1 camera deal with different challenges in depth map reconstruction. While light field cameras struggle to evaluate the third dimension in uniform areas, reflective or transparent surfaces raise difficulties for structured light cameras. In order to compare D images from Kinect V1 and Lytro camera, two measures are used: signal to noise ratio and precision.

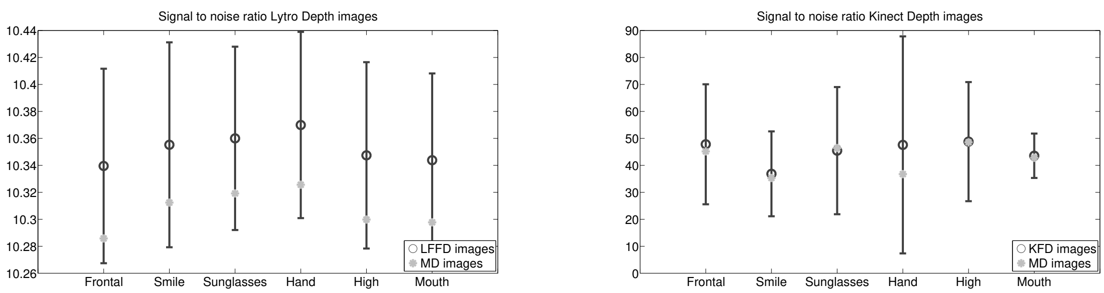
Signal to noise ratio on D images represents the measure of depth resolution. In order to evaluate the error on D map, each image is shifted and subtracted from itself. This operation is done 4 times, shifting on the four principal directions (right, left, up and down). The maximum range of values present in the image is considered as signal. For each picture, the standard deviation of the error distribution is compared with the signal.

Precision is related to image capability of reproducing the real depth value. The faceis assumed to be perfectly aligned with the camera and symmetric respect to a vertical plane passing through the center of the image. Thus, dividing a D picture along the symmetry axis, each pixel on the left side should have a correspondent value on the right. Signal



(a) (b)

Figure 1: Errorbar representing the distance distribution between LBP features representing the same subject. The graphics show how the statistics from JMD and LFFD (1a) and JMD and KFD (1b) are comparable for RGB images. The same experiment on depth map provides similar results.



(a) (b)

Figure 2: Errorbar representing the statistic of signal to noise ratio for Kinect images (2b) and Lytro images (2a). The mean value obtain from JMD images is always included in the confident interval estimated with LFFD and KFD databases .

to noise ratio has to be considered in this computation: if the resolution is low, precision is not significant. Each pixel lying on the left part of the image is compared with the symmetric one. If the difference between them is lower than the resolution, they are considered equal. Precision is defined as percentage of number of similar (with respect to the metric defined by signal to noise ratio) pixels between the left part and the right part of the face.

From one hand, signal to noise ratio results to be higher for images from Kinect with an average value of 47, while for images from Lytro is only 10. On the other hand, D images light field camera have a better precision (0.97 and 0.76 for Lytro and Kinect images respectively). The analysis paves the way to further studies dedicated to improving the acquisition method. The low signal to noise ratio is mostly caused by the narrow range of value in which face information are coded.

In fig 2 a measure of data variability is shown. The mean value and the standard deviation from LFFD and KFD are computed and a confident interval is represented. Mean value obtained from JMD results to be in the defined interval, proving that, even although LFFD and KFD present differences, the analysis done can be generalized.

2) Experiments: Compared with experiments conducted on RGB images, performances on D images are lower because the proposed algorithms are more suitable for textured images. Nevertheless, the recognition rate is still higher than random (tab I and tab II).

Experiment E1: D images from light field technology are largely influenced by illumination: this leads to significant low recognition rate for "High illumination" variation acquired with Lytro camera. Only 54% of images are well recognized while for structured light camera 92% of subjects is associated with the correct gallery data. In addition to time gap between first and second sessions, light variation influences the recognition rate on images obtained during the second session. Even though authors tried to reproduce the same conditions, light differences are present between sessions. All methods have lower performances if applied on face variations involving occlusions that change radically the depth of the image. In spite of more challenging images, results of analysis conducted on KFD data using LBP, LGBP and LBP3D-based methods are higher than 80% on "Smile" and "High illumination" variations of the first session. Light dissimilarities impact only slightly on the second session, where the recognition rate of "High illumination" stays

Table I: Percentage of rank-1 recognition rate for the first experiment (E1) on RGB, Depth images and fusion. Results are divided by face variation and session (S1 - S2).

E1		Smile (S1)	Sunglasses (S1)	Hand (S1)	High (S1)	Mouth (S1)	Neutral (S2)	Smile (S2)	Sunglasses (S2)	Hand (S2)	High (S2)	Mouth (S2)
Kinect	RGB	92.30	94.23	86.53	96.15	88.46	98.07	96.15	88.46	76.92	94.23	82.69
LBP	Depth	82.69	75	40.38	92.30	53.84	84.61	55.76	57.69	25	82.69	38.46
	Fusion	92,30	**96,15**	86,53	**98,07**	82,69	98,07	92,30	88,46	**80,76**	**96,15**	**84,61**
Lytro	RGB	100	81	86	97	91	88	82	51	56	79	50
LBP	Depth	80	48	30	46	62	25	22	14	10	22	16
	Fusion	100	**84**	**87**	97	**95**	88	82	**53**	**58**	**80**	**51**
Kinect	RGB	98.07	96.15	90.38	98.07	75	98.07	98.07	92.30	80.76	96.15	78.84
LGBP	Depth	82.69	57.69	38.46	78.84	55.76	65.38	55.76	40.38	26.92	75	34.61
	Fusion	92.30	96.15	86.53	98.07	**80.76**	98.07	92.30	82.69	80.76	96.15	**80.76**
Lytro	RGB	100	91	96	98	94	97	90	76	80	96	72
LGBP	Depth	86	45	45	54	66	29	26	15	13	28	19
	Fusion	100	**92**	96	98	**97**	97	89	74	80	**97**	**73**
Kinect	RGB	96.15	15.38	55.76	96.15	76.92	88.46	88.46	11.53	48.07	78.84	50
PCA	Depth	63.46	34.61	9.61	63.46	15.38	48.07	28.84	19.23	9.61	40.38	13.46
	Fusion	**98,07**	**44,23**	53,84	96,15	76,92	**92,30**	**90,38**	**30,76**	38,46	**82,69**	**57,69**
Lytro	RGB	97	49	52	94	68	81	68	28	32	66	36
PCA	Depth	73	25	18	43	45	13	17	8	5	14	7
	Fusion	**98**	**54**	**57**	94	69	80	68	**30**	**34**	66	**41**
Kinect	RGB	94.23	94.23	90.38	96.15	88.46	98.07	94.23	88.46	80.77	96.15	84.61
LBP3D	Depth	80.77	78.84	32.70	88.46	53.84	80.77	57.69	57.69	19.23	78.84	36.53
	Fusion	94.23	**96.15**	90.38	**98.07**	88.46	98.07	92.30	**90.38**	**82.69**	96.15	82.69
Lytro	RGB	100	80	88	97	90	89	82	53	56	79	51
LPB3D	Depth	80	46	31	47	63	25	23	15	9	23	17
	Fusion	100	**84**	87	97	**95**	88	82	53	**58**	80	51

Table II: Percentage of rank-1 recognition rate for the second experiment (E2) on RGB, Depth images and fusion of both. Results are divided by face variation and session (S1 - S2).

E2		Smile (S1)	Sunglasses (S1)	Hand (S1)	High (S1)	Mouth (S1)	Smile (S2)	Sunglasses (S2)	Hand (S2)	High (S2)	Mouth (S2)
Kinect	RGB	100	100	96.15	98.07	100	98.07	98.07	98.07	100	96.15
LBP	Depth	94.23	84.61	59.61	96.15	76.92	90.38	84.61	53.84	92.30	69.23
	Fusion	100	100	96,15	98,07	100	**100**	98,07	94,23	100	**100**
Lytro	RGB	100	96	92	99	95	100	92	88	99	93
LBP	Depth	82	44	29	39	58	86	31	26	33	58
	Fusion	100	96	**93**	99	**96**	100	92	**93**	**100**	93
Kinect	RGB	100	96.15	98.07	100	96.15	100	96.15	98.07	100	98.07
LGBP	Depth	80,76	63,46	36,53	88,46	59,61	84,61	61,53	46,15	84,61	53,84
	Fusion	100	**98.07**	92.30	98.07	**100**	100	**98.07**	90.38	100	**100**
Lytro	RGB	100	98	99	99	97	100	94	96	100	96
LGBP	Depth	84	47	32	54	59	92	38	36	52	61
	Fusion	100	98	99	99	**98**	100	94	**99**	100	96
Kinect	RGB	96.15	21.15	55.76	98.07	78.84	98.07	23.07	65.38	92.30	75
PCA	Depth	61.53	36.53	7.69	67.30	19.23	55.76	32.69	3.84	51.92	17.30
	Fusion	**100**	**48.07**	38.46	98.07	**80.76**	98.07	**40.38**	50	92.30	**76.92**
Lytro	RGB	97	48	56	93	64	100	49	62	98	60
PCA	Depth	73	25	16	42	44	79	17	11	37	31
	Fusion	**98**	**54**	55	93	**68**	100	**51**	62	**99**	**64**
Kinect	RGB	100	100	100	98.07	98.07	98.07	98.07	94.23	98.07	98.07
LBP3D	Depth	88.46	88.46	51.92	92.30	67.30	84.61	75	38.46	92.30	69.23
	Fusion	100	100	98.07	**100**	98.07	**100**	98.07	90.38	98.07	**100**
Lytro	RGB	100	96	92	99	97	100	91	88	99	93
LBP3D	Depth	83	43	28	37	59	86	32	27	34	58
	Fusion	100	96	93	99	96	100	91	**93**	100	93

higher than 75%.

Experiment E2: the percentage of images well recognized increases with respect to experiment E1, especially on pictures acquired during the second session. The improvement is obvious on "Smile" expression from LFFD, where the recognition rate increases up to 92% for LGBP-based method. LBP3D-based method does not outperform LBP-based method for D images. Previous studies [23], [24] proved that better results could be obtained analyzing different parts of the face separately. However, the main purpose here is to compare the impact of LBP3D features on D images from Lytro and Kinect. In both cases, the rank-1 recognition rates from LBP and LBP3D features are close. Recognition rate on D images using PCA-based method is low: percentage of well recognized subjects over all database is always lower than 35% for experiment E1 and than 40% for experiment E2.

The different performances obtained for LFFD and KFD images lead to conclude that D images from light field cameras are less informative for face recognition purposes than the one obtained with structured light sensors.

E. Fusion

The strength of both Kinect and Lytro camera is the double output, RGB and D image. For this reason, fusion between RGB and D images at score level is studied.

First, dissimilarity values (the scores) for both RGB and D images are normalized. In experiment E1, mean values and standard deviations for RGB and D images are computed from dissimilarities evaluated by comparing gallery data and "Neutral" expression recorded during the second session, while for experiment E2 the whole database is considered. Then, weights are fixed with a search grid, choosing the value that maximizes the recognition rate. Finally, Nearest Neighbor algorithm is evaluated on scores obtained in eq 1.

$$d_{fusion} = w * \hat{d}_{RGB} + (1 - w) * \hat{d}_{DM} \tag{1}$$

For both databases, fusion improves recognition. On KFD a gain over all face variation lower than 0.2% is achieved to be compared to a larger 1.6% for LFFD. Better results are not obtained by application of fusion on dissimilarities generated with LGBP features. The improvement obtained on LFFD images is lower than 0.5% and even negative when the method is applied to KFD images. This phenomenon could be explained by the small room for improvements left by the already high percentage of recognition on RGB images. As well as for RGB images, results obtained with LBP3D features are similar to the values achieved with LBP features. Analysis of PCA-based results is more interesting. The improvement due to fusion in experiment E1 is 7.8% and 3.2% respectively for KFD and LFFD images. Also for experiment E2 the gain is remarkable, 2% for both KFDatabase and LFFDatabase (tab I and tab II).

Experiment E1: for all databases and sessions, combining RGB and D images increases the percentage of well recognized subjects on "Sunglasses" and "Hand on face" variations when the method used is based on LBP features as well as on LBP3D features. The case of "High illumination" is of particular interest: results achieved with structured light camera are better than the ones obtained with light field technology, because of D map sensibility to the light. As previously mentioned, recognition rate does not change significantly for LGBP-based method. On the contrary, results obtained fusing RGB and D images using PCA-based method are particularly interesting. Recognition rate on RGB images regarding "Sunglasses" variation is low, especially on KFD images (15%). Fusion with D images triplicates the percentage of well recognized subjects for the first session and increase of 160% for the second. Improvements are also remarkable when the method is tested on LFFD.

Experiment E2: fusion between RGB and D images improves face recognition for both KFD and LFFD when of 3D occlusions occur. However, when the face variation considered does not present depth difference from training data (like "Smile" variation), the impact of fusion is less important. Moreover, the high recognition rate for RGB images gives a small room for improvements.

F. Results consistency

The tests conducted on LFFD and KFD need to be generalized in order to prove that small variation of the acquisition environment or the identity of represented subjects do not deeply influence the results.

With this aim, tests on JMD have been conducted. While LFFD and KFD have been collected in different periods and represent different subjects, the JMD includes images of the same subject in the same environment acquired at the same time with both the sensors. Mean and standard deviation of distances between samples from the same subject, respectively from LFFD and KFD images, are computed with the aim of creating a confidence interval. In fig 1 is shown how the average value of the distances between samples from the same subject from JMD fall in the interval defined previously for each face variation for both Lytro and Kinect images for LBP-based method. The same results are obtained for LGBP and LBP3D-based methods. The procedure could not be followed for PCA-based methods. The space where the distance computed is dependent on the database. The same analysis is computed on D images with similar results, proving that the analysis is not influenced by the differences between the databases.

IV. Conclusion

In this work, the potential for face recognition of two particular technologies, structured light and light field cameras, has been investigated. Tests on IST-EURECOM Light Field Face Database and on EURECOM Kinect Face Database have been done, taking advantages of their similar protocol acquisition and data structure. Two experiments have been set up, one using as gallery one picture per subject and one keeping two images per person, the former from the first session, the latter from the second. The analysis has shown how the time gap between the sessions could be really influential in face recognition. Tested algorithms have been proven more suitable for D images acquired with structured light sensor that light field camera. Particular relevant is the fact that D images from light field cameras are strongly affected by illumination. An home made database of 20 persons has been acquired with both Lytro Illum camera and Kinect sensor at the same time in order to consolidate the different impact of these technologies on face recognition.

RGB and D images have been fused at score level verifying the improvement of recognition ratio, especially when the image presents tridimensional occlusion like in the case of "Sunglasses" variation. Signal to noise ratio on light field D images appears strongly influenced by the acquisition process. The information stored in depth map from LFFD do not take all range of possible eligible values. For this reason, understanding how to improve the quality of Lytro image of face data acquisition would be an essential topic still to be discussed. The comparison of Lytro and Kinect images illustrates the benefits generated by either one or the other technology and reveals the drawback of their use. This paper could be considered as baseline whenever there is the necessity to choose one of the studied sensors.

Even if in this work hand-crafted features have been used, this result may be useful in further analysis based on learned features. A natural development of this work could be verifying the consistency of the results when using deep-learning algorithms and to test the proposed procedure on bigger databases illustrating different acquisition conditions.

V. Acknowledge

This work has been partially supported by European Union's Horizon 2020 research and innovation program under grant agreement No 700259, (PROTECT).

References

[1] R. Min, N. Kose, and J. Dugelay, "Kinectfacedb: a kinect database for face recognition," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 07 2014.

[2] R. Raghavendra, K. B. Raja, and C. Busch, "Presentation attack detection for face recognition using light field camera," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 1060–1075, March 2015.

[3] M. Hossny, D. Filippidis, W. Abdelrahman, H. Zhou, M. Fielding, J. Mullins, L. Wei, D. Creighton, V. Puri, and S. Nahavandi, "Low cost multimodal facial recognition via kinect sensors," in *Land Warfare Conference 2012*, 2012.

[4] R. I. Hg, P. Jasek, C. Rofidal, K. Nasrollahi, T. B. Moeslund, and G. Tranchet, "An rgb-d database using microsoft's kinect for windows for face detection," in *2012 Eighth International Conference on Signal Image Technology and Internet Based Systems*, Nov 2012, pp. 42–46.

[5] B. Li, A. S. Mian, W. Liu, and A. Krishna, "Using kinect for face recognition under varying poses, expressions, illumination and disguise," in *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, Jan 2013, pp. 186–192.

[6] G. Goswami, S. Bharadwaj, M. Vatsa, and R. Singh, "On rgb-d face recognition using kinect," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Sept 2013, pp. 1–6.

[7] T. Huynh, R. Min, and J. Dugelay, *An Efficient LBP-Based Descriptor for Facial Depth Images Applied to Gender Recognition Using RGB-D Face Data*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 133–145.

[8] N. Erdogmus and S. Marcel, "Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Sept 2013, pp. 1–6.

[9] R. Raghavendra, K. B. Raja, B. Yang, and C. Busch, "Guclf: a new light field face database," pp. 8785DB–8785DB–6, 2013.

[10] R. Raghavendra and C. Busch, "Presentation attack detection on visible spectrum iris recognition by exploring inherent characteristics of light field camera," in *IEEE International Joint Conference on Biometrics*, Sept 2014, pp. 1–8.

[11] S. Kim, Y. Ban, and S. Lee, "Face liveness detection using a light field camera," *Sensors*, vol. 14, no. 12, pp. 22 471–22 499, 2014.

[12] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recogn. Lett.*, vol. 24, no. 13, pp. 2115–2125, Sep. 2003. [Online]. Available: http://dx.doi.org/10.1016/S0167-8655(03)00079-5

[13] M. L. Gavrilova and M. M. Monwar, *Current Trends in Multimodal Biometric System—Rank Level Fusion*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 657–673. [Online]. Available: https://doi.org/10.1007/978-3-642-22407-2_25

[14] S. C. Dass, K. Nandakumar, and A. K. Jain, "A principled approach to score level fusion in multimodal biometric systems," in *Audio- and Video-Based Biometric Person Authentication*, T. Kanade, A. Jain, and N. K. Ratha, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 1049–1058.

[15] M. He, S.-J. Horng, P. Fan, R.-S. Run, R.-J. Chen, J.-L. Lai, M. K. Khan, and K. O. Sentosa, "Performance evaluation of score level fusion in multimodal biometric systems," *Pattern Recogn.*, vol. 43, no. 5, pp. 1789–1800, May 2010. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2009.11.018

[16] K. C. K. W. Bowyer and P. Flynn, "A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition," *Computer vision and image understanding*, vol. 101, pp. 1–15, 2006.

[17] A. Sepas-Moghaddam, V. Chiesa, P. Lobato Correia, F. Pereira, and J. Dugelay, "The ist-eurecom light field face database," in *IWBF 2017, 5th International Workshop on Biometrics and Forensics, 4-5 April 2017, Coventry, UK*, Coventry, ROYAUME-UNI, 04 2017.

[18] M. Feng, S. Zulqarnain Gilani, Y. Wang, and A. Mian, "3D Face Reconstruction from Light Field Images: A Model-free Approach," *ArXiv e-prints*, Nov. 2017.

[19] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Dec. 2009.

[20] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Jan. 1991.

[21] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, Dec 2006.

[22] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 1, Oct 2005, pp. 786–791 Vol. 1.

[23] Y. W. Y. Huang and T. Tan, "Combining statistics of geometrical and correlative features for 3d face recognition," in *Proceedings of the British Machine Vision Conference (BMVA)*, 2006, pp. 90.1–90.10.

[24] J. B. C. Neto and A. N. Marana, "Face recognition using 3dlbp method applied to depth maps obtained from kinect sensors," in *Workshop de Visao Computacional (WVC)*, 2014.