

On Multi-View Face Recognition Using Lytro Images

Valeria Chiesa
EURECOM

Biot Sophia Antipolis, FRANCE
valeria.chiesa@eurecom.fr

Jean-Luc Dugelay
EURECOM

Biot Sophia Antipolis, FRANCE
Jean-Luc.Dugelay@eurecom.fr

Abstract—In this work, a simple and efficient approach for recognizing faces from light field images, notably from Lytro Illum camera, is proposed. The suggested method is based on light field images property of being rendered through a multi-view representation. In the preliminary analysis, feature vectors extracted from different views of the same Lytro picture are proved different enough to provide complementary information beneficial for face recognition purpose. Starting from a set of multiple views for each data, face verification problem is tackled and results are compared with those achieved with classical 2D images simulated using a single view, i.e. the central one. Two experiments are described and, in both cases, the presented method shows superior performances than standard algorithms adopted by classical imaging sensors.

Index Terms—Multi-view, light field images, face recognition, Lytro camera

I. INTRODUCTION

In the last decade, face recognition has become important in several domains, improving public security (e.g. in border controls) or daily live quality (e.g. bank security accesses or security key on smart devices). The ability to automatically recognize the identity of a person is deeply influenced by the acquisition technology. Several studies have been done in order to adapt existent algorithms on different kind of data or to develop new methods customized for specific images ([1], [2]).

Hand-crafted features have been revealed efficient for long time [3]. Recently, the research has moved in machine learning direction and new studies ([4], [5]) have proven how neural network algorithms outperform hand-crafted features-based methods on pattern recognition and in particular on face recognition and verification. However, face recognition based on light field images has been little investigated because of the scarcity of open-access databases.

In [6], the first light field face database is presented. Authors create and test an algorithm able to detect the best focused image among all the possible focus levels rendered by a single Lytro image. In [7] and [8], light field images have been also used for video-surveillance tasks. In the former, authors present a super-resolution scheme to enhance high frequencies and to improve face recognition, while in the latter, a weight assessment scheme is described with the aim of fusing different images. Light field cameras have revealed

beneficial also in presentation attack detection ([9], [10], [11]).

The improvements achieved by using multi-view data on face recognition have already been studied. Usually, views are collected with different devices at the same time [12] or with the same sensor with different shots [13]. In both cases, the data acquisition can be complex or require high degree of cooperation from the subject. Similar studies are presented in [14] where authors tackle face recognition problem creating multi-view representation from RGB-D images collected with Kinect sensor. Data are processed with a deep-learning algorithm in order to investigate how view-space partitioning impacts on face recognition performance. Also in [15] a novel approach based on multi-view properties is used to recognize subjects illustrated in different poses. Synthetic face images are generated to imitate the other pose variations, thus helping in the recognition process under different perspectives.

A multi-view based approach for light field technology has been barely studied in [16]. Although if the variation in view representation is small, Lytro images have some advantages related to the collection modality: in fact, with a single shot, several views are acquired simultaneously, without any camera alignment problems. The goal of this work is to investigate the presence of complementary information in Lytro multi-view representation beneficial for face recognition.

The main contributions of this article are:

- proving that multi-view representation of light field images, recorded by a Lytro Illum camera, provides complementary information beneficial for face recognition;
- analyzing the relation between view shifting and recognition algorithms;
- suggesting a simple and efficient approach to exploit multi-view representation potentiality on face recognition;
- illustrating the improvements of the proposed method over the classical 2D RGB images approaches;

In the section II light field cameras are briefly described, focusing on the Lytro Illum camera. Database, preprocessing and features used in this work are shown in section III, IV and V. Preliminary analyses are explained in section VI and new metrics are defined in section VII. In section VIII two experiments are described and results are commented. Finally, conclusions are discussed in section IX.

II. LIGHT FIELD CAMERAS

Light field camera can be structured like cameras gantry or a single automated movement camera restricted to still object imaging. The camera used in this work (Lytro Illum) is composed by a standard sensor sensible to visible frequencies, one main lens and a microlens array and it is sold by Lytro, Inc. [17] to large audience. Light field cameras are based on the concept that the entire visible space can be described with a function dependent of 7 parameters corresponding to time, light frequency, coordinates of point of view and position of the point observed. A more extensive explanation could be found in [18]. This kind of cameras can be used outdoors and have no proximity limitations.

The size of the rendered images can reach 2450x1634 pixels. The CMOS image sensor of Illum Lytro camera measures 1/2" (6.4 x 4.8 mm) and the sensitivity can be varied from ISO 80 to 3200. The main lens has a size of 30-250mm and the optical zoom of 8.3x. It features f/2.0 aperture with 1:3 macro capability.

Images can be rendered in several ways, including multi-view representation of the scene, each view defined as the picture of the scene collected with a slightly different yaw and pitch angles. Therefore, the transformation considered is not a simple bidimensional shifting but the difference between views involves depth dimension, exploiting the 3D properties of light field cameras.

Standard 2D images are easily achievable from light field data: each view can be considered as classical RGB picture independent from others.

III. DATABASE

As far as we know, only two face databases collected with light field camera are publically available. The first, "GUCLF: a new light field face database" [6], has been published in 2013 and it contains post-processed images acquired in a controlled environment with both standard and light field cameras.

The second, "Light Field Face Database" (LFFD) [19], has been published in 2017. The database consists in a set of 100 subjects with 20 face variations, including different illuminations, poses, occlusions and expressions. All 20 modalities have been collected during two sessions separated by at least one month. In contrast to GUCLF, raw images and calibration data of LFFD are available to the end user. For this reason, the proposed work is carried on LFFD images. The analysis has been restricted to six face variations: "Frontal face" expression (FF), "Happy" expression (HE), "Open mouth" expression (OM), "Sunglasses" occlusion (SO), "Hand on mouth" occlusion (HM) and "High" illumination (HI) for a total of 1200 images.

IV. PRE-PROCESSING

Raw data are processed with Lytro Power tool [20], a free software provided by Lytro, Inc. This software is able to manipulate raw Lytro data and to render light field images as multi-view RGB images, each one collected from a slightly shifted point of view. For this work, each data is transformed

in a 5x5 RGB view matrix, each view with size 2022x1404 pixels (Fig 1). Unfortunately the algorithm used to obtained the multi-view representation from the raw data is not provided by Lytro, Inc.



Fig. 1: Example of multi-view representation of Lytro image

Each view is initially processed as single picture. The face represented is detected, aligned and cropped with a pre-trained model based on HOG features available on the free library DLIB [21].

V. FEATURES

Face analysis is carried on using 3 sets of features. Two of them are classic hand-crafted features (LBP [22] and LGBP [23]), the latter are based on deep-learning algorithm (OpenFace [24]). All of them are selected due to their good performances on face recognition. LBP features are extracted with a well known-method computed on blocks of 8x8 pixels [22]. The same algorithm is used to find LGBP features from Gabor filtered images [23].

OpenFace (OF) [24] is based on neural networks algorithm [5] and implemented in Python and Torch. This algorithm is designed to map faces to a 128-dimensional Euclidean hypersphere so that the Euclidean distance is able to discriminate different subjects, providing excellent recognition performance on several datasets. This property allows to successfully use a simple classifier based on Euclidean distance in 128-dimensional space to discriminate between matching and mismatching samples.

A simple test on the database is carried on LFFD in order to evaluate features performances. Each light field data is associated with classical 2D RGB image considering only the central view. OF, LBP and LGBP features are extracted and a classifier for each method is created using as reference data the *frontal face* images collected during the first session and as probe set all other face variations. Results are shown in tab I.

Baseline analysis shows how recognition achieved with OF features outperforms the results obtained with LBP and LGBP features. That result has already been proved [25] and the comparison of handcrafted features and deep-learning methods is not the main purpose of this work.

VI. PRELIMINARY ANALYSIS

The 25 views of the same plenoptic image are only slightly different (Fig 1). Before developing a new method able to

	OF	LBP	LGBP
EER	0.0156	0.2099	0.1760
FMR1000	0.0938	0.7692	0.5623
ZeroFMR	0.2981	0.8773	0.7271

TABLE I: EER, FMR1000 and ZeroFMR related to OF, LBP and LGBP-based methods on classical 2D images obtained from LFFDatabase. OF-based method outperforms handcrafted algorithms.

exploit multi-view properties, it is necessary to study how the change of perspective impacts on features computation. With this aim, OF, LBP and LGBP features are computed on all views of the same raw data separately, as if they were independent 2D pictures. Distances among feature vectors of central view and all other views considered as classical RGB images are computed.

In tab II statistics related to the Relative Standard Deviation (RSD) are reported. Mean value and standard deviation are evaluated on all views of the same raw data and successively summarized on the whole considered database to be shown. In this preliminary step, all modalities and sessions of LFFD are used indifferently because intra-image differences (i.e. differences between views from the same plenoptic image) are not influenced by face variation.

inter image	intra images	OF	LBP	LGBP
RSD	mean	0.4411	0.2858	0.2975
	var	0.0019	0.0434	0.04

TABLE II: Relative Standard Deviation statistics: the average value obtained with OF features is higher than the others. Therefore, a stronger impact of multi-view representation when using OF-based algorithm is expected.

The higher value of the RSD related to OF features indicates a stronger average variance between views of the same plenoptic image when they are represented by these features. This consideration suggests better results on multi-view fusion if performed with OF features.

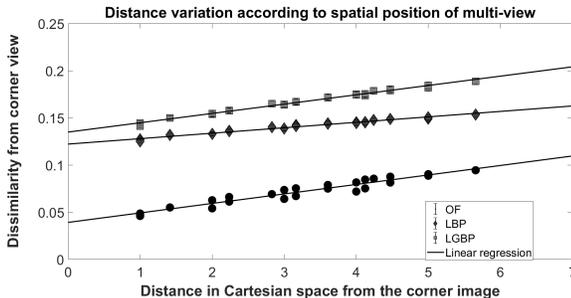


Fig. 2: Normalized Euclidean distance of each view respect to corner view in the same image v.s. space distance between the considered views. The solid lines are linear fit to the data, showing a linear relation between view shift and recognition algorithms.

In order to study the maximum variation achievable from a single data, we compute the distance between the corner view and all other views. View shifting is not dependent on the direction (horizontal or vertical). Thus, pitch and yaw angles are considered equally. The average normalized dissimilarity scores achieved in features space are evaluated as a function of the view shift. View shift is calculated as $D = \sqrt[2]{(i-1) + (j-1)}$ where i and j are the view index in the pictures array, so that the central view will be at $D = 2$. The linear trend between distances in features space and view shift is illustrated in Fig 2. Better performances of OF features based methods respect to LBP or LGBP are expected because of the resulting steeper slope. Preliminary analyses show that the information stored in a multi-view representation of a human face are richer than a standard RGB image, especially when faces are mapped in the 128-dimensional hypersphere defined by OF features. Thus, further analyses are carried on in parallel with all described feature extraction algorithms but, since the conclusions are similar for all methods, only results related to OF features are commented.

VII. DISTANCES

In order to use the additional information stored in light field data, feature vectors are extracted from all views of the same image. Usually, classical algorithms compare two images using one feature vector for each data. The vectors are processed to assess if the two pictures represent the same subject. Since light field data provides multiple views of the same subject, an additional step is computed: feature vectors are extracted from all views attached to raw data and, then, cross distances among all views of the two images are computed. The multiple values are reduced down to a single one for making a decision with the proposed functions. Eight simple distances are described.

Let A and B be the set of feature vectors describing the views from two images and d_e the Euclidean distance between elements in a set, it is possible to define:

- Baseline: distance between central view of two images.
- Min distance: minimum value through all the possible cross distances. $d_{min} = \min_{a \in A, b \in B} (d_e(a, b))$
- Min distance corners: the minimum value through all the possible distances between corner views. $d_{minc} = \min_{a \in A_c, b \in B_c} (d_e(a, b))$
- Mean distance: the average value of all the possible cross distances. $d_{mean} = \text{mean}(d_e(a, b) \quad \forall a \in A, \forall b \in B)$
- Mean distance corners: average value through all possible distances between corner views. $d_{meanc} = \text{mean}(d_e(a, b) \quad \forall a \in A_c, \forall b \in B_c)$
- Max distance: the maximum value through all the possible cross distances. $d_{max} = \max_{a \in A, b \in B} (d_e(a, b))$
- Max distance corners: the maximum value through all the possible distances between corner views. $d_{maxc} = \max_{a \in A_c, b \in B_c} (d_e(a, b))$
- Hausdorff mean distance: $d_{Hmean} = \frac{1}{\#A + \#B} \{ \sum_{a \in A} \min_{b \in B} d_e(a, b) + \sum_{b \in B} \min_{a \in A} d_e(a, b) \}$
- Hausdorff max distance: $d_{Hmax} = \text{Max} \{ \max_{a \in A} \min_{b \in B} d_e(a, B), \max_{b \in B} \min_{a \in A} d_e(b, A) \}$

Both d_{min} and d_{max} distances are studied. In fact, considering the minimum value of dissimilarity, the distance between matching samples decreases as well as distance between mismatching samples. Vice-versa, when d_{max} distance is applied, the dissimilarity increases. In the first case, the probability of false matching increases, in the second case, the probability of false non-matching raises. Distances evaluated using only corner views are studied in order to consider algorithms computationally less expensive.

VIII. EXPERIMENTS

Two verification experiments are set up. The first follows a closed-set protocol where each subject considered during the test phase is also used for validation. The second is an open-set experiment where the system is created with samples of 80 subjects and tested on the remaining 20. A cross-validation model is applied in order to generalize the conclusions.

In both cases, *frontal face* images from the first session of LFFD are used as reference data. Distances are computed between each plenoptic image and each reference data. During the validation phase, classifiers are defined (notably the distance representing the threshold below which the probe sample is matched with the reference one). Performances are shown comparing False Acceptance Rate and False Rejected Rate evaluated on the test set.

A. Closed-set experiment

In this experiment, we investigate how the suggested technique can be applied to face verification over different time span. Thus, we select validation and test sets from the different acquisition sessions of the database.

Validation phase: Images acquired during the first session of LFFD are used. *Frontal face* variation is considered as reference (one image for each subject for a total of 100 raw data) and all other variations from the first session are used as validation set (one image for each subject for each modality for a total of 500 raw data). Features distances between all reference and validation samples are computed in order to define nine (eight proposed and baseline) pools of linear classifiers, one for each distance definition. From each pool, the classifier corresponding to equal error rate (EER) is chosen.

Test phase: The test set is composed of all data from the second session considered in this work (600 raw data).

Results: In Fig 3 the False Acceptance Rate (FAR) vs False Rejection Rate (FRR) on the test set of the nine (eight proposed and baseline) classifiers created with the different distances from OF feature vectors is represented. While the classifier based on d_{min} distance obtains low FRR in spite of high FAR, the one with d_{max} distance shows the opposite results. Both d_{mean} and d_{Hmean} classifiers outperform baseline classifier obtaining accuracy respectively equal to 99.20% and 99.13% v.s. 98.65%, notwithstanding the fact that the latter already perform at an high level of accuracy.

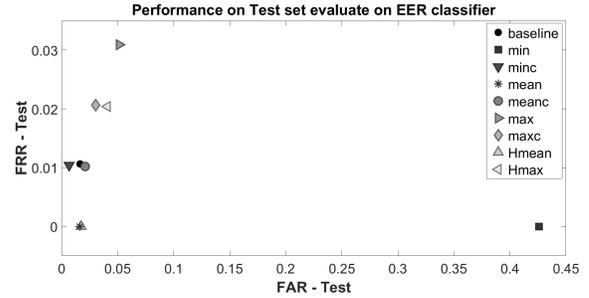


Fig. 3: Ex 1: Performance of classifiers evaluated on test set represented as False Acceptance Rate (FAR) vs False Rejected Rate (FRR). The best performances are obtained with d_{mean} and d_{Hmean} distances

B. Open-set experiment

The purpose of the open-set experiment is to demonstrate how the described algorithms can be successfully tested on subjects that are not considered during the validation phase.

Validation phase: Only 80% of LFFDatabase subjects are considered during the validation process. All face variations of the 80 subjects are used to define the classification threshold (880 raw data). A pool of classifiers for each distance is created and the one corresponding to EER is chosen.

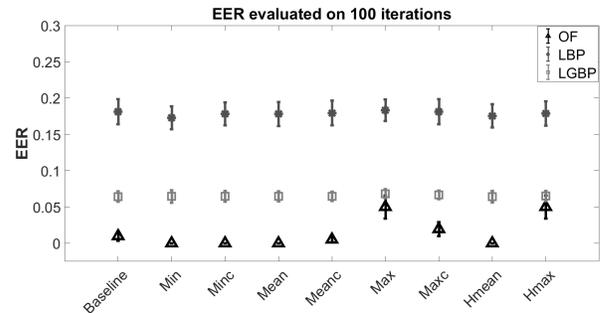


Fig. 4: Ex 2: EER evaluated on the validation set for different distances and features. While EERs relative to LBP and LGBP features have a stable behavior among different distances, EERs obtained with OF features classifiers present a higher variance

Test phase: All face variations illustrating the remaining subjects are included in the test set (220 raw data).

Cross-validation: With the aim of improving analysis stability, a cross-validation algorithm is applied. Validation and test phases are repeated 100 times splitting the database randomly. In Fig 4 a representation of EER distributions for OF, LBP and LGBP evaluated on validation set is shown. As suggested in section VI, OF features have not only lower EER but also a higher variance among different distance classifiers.

Results: In Fig 5 results obtained with OF features are illustrated. As for the first experiment, d_{min} and d_{minc} classifiers outperform baseline with a respective accuracy of 99.80% and 99.78% v.s. 99.27%. d_{max} , d_{Hmax} and d_{maxc} classifiers do not improve verification results.

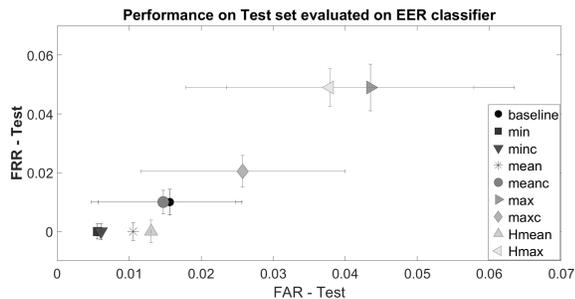


Fig. 5: Ex 2: Performance of classifiers evaluated on test set represented as False Acceptance Rate (FAR) vs False Rejected Rate (FRR). The best performances are obtained with d_{min} and d_{minc} distances

IX. CONCLUSION

This work tackles face recognition problem with light field images, by using LFFD database. The light-field acquisition are rendered as a 5 by 5 multi-view images for data analysis.

We study how light field images are richer than standard RGB 2D-picture and we propose a method to use the extra-information to improve face recognition algorithms.

Cross distances among views of two images are computed. Eight new distances customized for reducing down multiple values to a single one are defined. Two verification experiments are carried on. The close-set experiment shows how d_{mean} and d_{Hmean} distances can improve face verification performances on images collected in a different period than the samples used during the validation phase. The open-set experiment demonstrates how d_{min} and d_{minc} distances could be successfully tested on subjects that are not considered during the validation phase.

The analysis paves the way to a more exhaustive study on impact of multi-view fusion on face recognition and verification when bigger pitch and yaw angles are applied between subject face axes and camera. This can be achieved with Lytro Illum by taking closer snapshots or with different kind of light field cameras (e.g. Raytrix).

X. ACKNOWLEDGEMENTS

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

REFERENCES

- [1] 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.
- [2] 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.
- [3] ivind Due Trier, Anil K. Jain, and Torfinn Taxt, "Feature extraction methods for character recognition-a survey," *Pattern Recognition*, vol. 29, no. 4, pp. 641 – 662, 1996.
- [4] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.

- [5] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," pp. 815–823, 06 2015.
- [6] R. Raghavendra, Kiran B. Raja, Bian Yang, and Christoph Busch, "Guclf: a new light field face database," 2013.
- [7] R. Raghavendra, K. B. Raja, B. Yang, and C. Busch, "Comparative evaluation of super-resolution techniques for multi-face recognition using light-field camera," in *2013 18th International Conference on Digital Signal Processing (DSP)*, July 2013, pp. 1–6.
- [8] R. Raghavendra, Kiran B. Raja, Bian Yang, and Christoph Busch, "Improved face recognition at a distance using light field camera & super resolution schemes," in *Proceedings of the 6th International Conference on Security of Information and Networks*, New York, NY, USA, 2013, SIN '13, pp. 413–416, ACM.
- [9] S. Kim, Y. Ban, and S. Lee, "Face liveness detection using a light field camera," *Sensors*, vol. 14, no. 12, pp. 22471–22499, 2014.
- [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] A. Sepas-Moghaddam, L. Malhadas, P. L. Correia, and F. Pereira, "Face spoofing detection using a light field imaging framework," *IET Biometrics*, vol. 7, no. 1, pp. 39–48, 2018.
- [12] M. Y. Shams, A. S. Tolba, and S. H. Sarhan, "A vision system for multi-view face recognition," *CoRR*, vol. abs/1706.00510, 2017.
- [13] Zhenyao Zhu, P Luo, X Wang, and X Tang, "Multi-view perceptron: A deep model for learning face identity and view representations," vol. 1, pp. 217–225, 01 2014.
- [14] Donghun Kim, Bharath Comandur, Henry Medeiros, Noha M. Elfiky, and Avinash C. Kak, "Multi-view face recognition from single rgbd models of the faces," *Computer Vision and Image Understanding*, vol. 160, pp. 114 – 132, 2017.
- [15] Koichiro Niinuma, Hu Han, and Anil K. Jain, "Automatic multi-view face recognition via 3d model based pose regularization," 09 2013.
- [16] A. Moghaddam, P.L. Correia, and F. Pereira, "Light field local binary patterns description for face recognition," in *IEEE International Conf. on Image Processing - ICIP*, September 2017, pp. –.
- [17] Lytro Inc, "Lytro website," 2017.
- [18] R. Ng, *Digital Light Field Photography*, Ph.D. thesis, Stanford, CA, USA, 2006, AAI3219345.
- [19] A. Sepas-Moghaddam, V. Chiesa, P. Lobato Correia, F. Pereira, and J.L. 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.
- [20] Inc. Lytro, "Lytro power tool," .
- [21] Davis E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Dec. 2009.
- [22] 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.
- [23] 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*, Oct 2005, vol. 1, pp. 786–791 Vol. 1.
- [24] Brandon Amos, Bartosz Ludwiczuk, and Mahadev Satyanarayanan, "Openface: A general-purpose face recognition library with mobile applications," Tech. Rep., CMU-CS-16-118, CMU School of Computer Science, 2016.
- [25] S. Setiowati, Zulfanahri, E. L. Franita, and I. Ardiyanto, "A review of optimization method in face recognition: Comparison deep learning and non-deep learning methods," in *2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Oct 2017, pp. 1–6.