LFHOG: A DISCRIMINATIVE DESCRIPTOR FOR LIVE FACE DETECTION FROM LIGHT FIELD IMAGE

Zhe Ji, Hao Zhu, Qing Wang

School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China

ABSTRACT

How to avoid the invading of the attack in the biometric system, such as 2D printed photos, gradually becomes an important research hotspot. In this paper, we present a novel descriptor in light field to tackle the issue. Based on the angular and spatial information in light field, the proposed light field histogram of gradient (LFHoG) descriptor is derived from three directions, including vertical, horizontal and depth. Different with traditional HoG in 2D image, the gradient in depth direction is distinctive in light field. To validate the effectiveness of the proposed LFHoG descriptor, experiments have been carried out on light field datasets taken by a Lytro camera. The descriptor can achieve 99.75% accuracy on the user collected dataset, which proves the correctness and effectiveness of the LFHoG descriptor.

Index Terms— Light field, Light field histogram of gradient (LFHoG), Live face, Face detection

1. INTRODUCTION

The light field imaging techniques [1] have been developed quickly in recent years and commercial light field cameras, such as Lytro and Raytrix have been available in the market. The reason why light field imaging is popular is that it can collect more angular information for an incident ray while traditional imaging only records one angular information. As a result, the depth estimation becomes very easy and it makes many applications possible, e.g. image super-resolution [2], refocusing [3] and 3D reconstruction [4]. The light field can also be utilized in the biometric system [5] to enhance face/iris recognition by its abundant information and it is becoming a developing tendency in the biometric system to equip light field cameras.

Nowadays, the face often works as a characteristic on the identification or verification system. The weakness that machine cannot distinguish whether the face is live or not becomes the biggest loophole of the identification system. It often happens that some intruders invade identification systems with the forged faces. For this reason, the technique of live face detection is necessary for the identification system and it is essential to defense against the invasion and baleful attack.

To complete the function of live face detection, lots of techniques have been developed in the last few years. There are some techniques that utilize the characteristics of recapture 2D images, such as the lost sharpness and detail [5, 6], difference in light distribution [7, 8] and motionless [9, 10] and so on. But these are all weak clues and only limited to specific cases which are easily influenced by the environment, so these characteristics cannot distinguish the live face and recaptured 2D photo very well. Besides, there are also many methods based on the 3D facial model of human [11], but the acquisition of accurate 3D facial model of human is expensive and hard. In addition, 3D facial models only reserve the structure information of the face but lose the texture information. Fortunately, the emerging light field technology may overcome these disadvantages in live face detection.

There are some works trying to detect the live face in light field, but they all follow the straightforward strategy that first detects the face area and then judges whether it is a live face. Kim et al. [12] proposed detecting the face in a sub-image by using a LBP descriptor and then analyze the edge and ray difference in 4D light field to detect the live face. Raghavendra et al. [13] refocus the light field in different depth, then they detect the face and evaluate the focusness of each refocused image, finally the live face is distinguished by analyzing the variation of focusness. Ghasemi et al. [14] propose to extract energy feature vector from EPIs (Epipolar Plane Image) to distinguish the flat surface from non-flat surface, but they only propose the method to judge the face instead of a method to detect the face.

Differently from the previous work in light field face detection, we build the light field descriptor which comes from the light field raw data. And live face and the printed face are detected and distinguished by a multiple classifier directly. The pipeline of our approach is shown in Fig.1.

Considering the special imaging process of light field images, we define our gradient in depth direction and propose a light field HoG descriptor (LFHoG) which includes the gradients in three directions (vertical, horizontal, and depth). Besides we represent the LFHoG in a spherical coordinates system and use it as a feature which only exists in light field.

The work is supported by NSFC funds (61272287, 61531014) and research grant of State Key Laboratory of Virtual Reality Technology and Systems (BUAAVR-15KF-10).

Then we use a linear SVM classifier to train the database and the experiment results show LFHoG descriptor can achieve an encouraging performance.

Our contributions are summarized as follow:

1) We propose a novel LFHoG descriptor in light field, which integrates the distribution of color intensity and the distribution of scene depth simultaneously.
2) Based on the LFHoG descriptor, we can realize the live face detection in one step, and do not need to first detect the face area and then distinguish its liveness.

2. LIGHT FIELD HOG DESCRIPTOR

Similar to the traditional HoG in 2D image, the histogram of gradient also exists in light field with a different style. In this section, we will explore a novel HoG, LFHoG, in light field. Benefiting from the abundant angular information of an incident ray, the depth distribution of a scene can be easily obtained in light field. Hence in addition to the structure information which can be obtained in 2D images, the depth information of a scene also should be considered in the LFHoG. The LFHoG is the just the histogram of gradients in light field, where is composed of gradients in vertical, horizontal and depth directions. Since the gradient along depth direction is not simply same as the others along vertical and horizontal directions, it brings up a new issue how we define the gradients in light field.

Apparently, the edge and texture of objects are important features to describe the appearance and shape of objects since they are well described by the distribution of gradient direction in 2D image. The traditional HoG [15] just utilizes the human perception for object detection. Therefore the gradient of LFHoG in vertical and horizontal directions can be defined as the gradient of an image in light field.

The gradient along depth direction is a new component in LFHoG, which is also the representation of the distribution of the scene. The policy of the gradient form is tightly related with the point spread function in camera model. The point spread function (PSF) [16] is the space-invariant spatial response of an imaging system to a point light source, which often acts as a low-pass filter. Since a lowpass filter destroys small details, the sharpness of the detail is connected tightly with the radius of PSF. Therefore, we can get the message that the radius of PSF in a real scene is changing with the focal depth but it keeps consistent in a flat scene. In order to reflect the difference in different focal depth, the gradient in depth direction can be defined as the difference between the focal planes with distinctive focal depth.

Based on the analyses of the gradients of three directions in LFHoG, it is obvious that the descriptor has integrated the distribution of the color intensity and depth distribution of a scene. On account of the LFHoG, the live face detection became an easy process. Then the form of LFHoG will be introduced in detail at the process of live face detection.

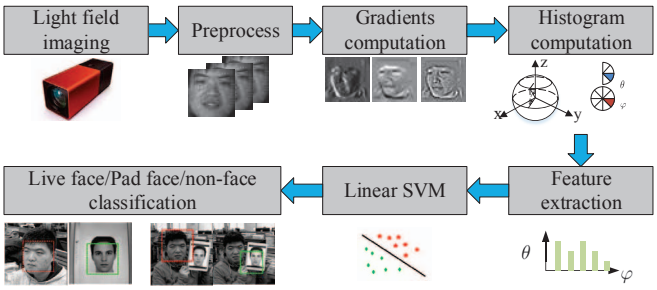


Fig. 1. The pipeline of our LFHoG extraction and live face detection in light field.

The process of live face detection with LFHoG can be divided into several steps as follow, as shown in Fig.1.

Preprocessing. Since the face may be lying in different depth when the plenoptic image is obtained, the light field ought to be refocused to the corresponding depth. The depth map can be obtained by structure tensor [1]. The next step is refocusing the light field on the depth where the nose is in. An illumination of light field representation in 2D case is as follows,

$$L_\alpha(x,u) = L\left(x + u\left(1 - \frac{1}{\alpha}\right), u\right) \tag{1}$$

where $L(x,u)$ is the 2D representation of light field, x is one of the spatial dimensions and u is one of the angular dimensions. More importantly, α is related to the depth of the scene.

Gradient Computation. After the preprocessing of light field, it has been rectified in the reference light field. Then the gradients of three directions can be computed. The gradients in vertical and horizontal directions are same as the traditional HoG, i.e. and in horizontal and vertical directions in the rendered image [2] of the reference light field. In accordance with the analysis of the gradient in depth direction, it can be calculated by the Eq.2,

$$dz = E_{\alpha_1}(x,y) - E_{\alpha_2}(x,y) \tag{2}$$

where $E_{\alpha_i}(x,y)$ are the refocused images in the reference depth α_1 and α_2 respectively, which means that $\alpha_1 = 1.0$ and α_2 can be changing with our setups. The $E_\alpha(x,y)$ can be computed by the Eq.3:

$$E_\alpha(x,y) = \frac{1}{\alpha^2}\iint L\left(u\left(1-\frac{1}{\alpha}\right)+\frac{x}{\alpha}, v\left(1-\frac{1}{\alpha}\right)+\frac{y}{\alpha}\right) dudv. \tag{3}$$

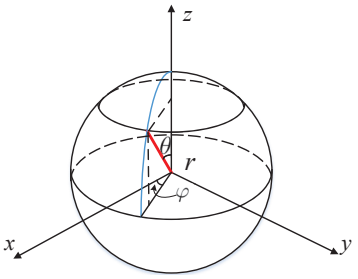
Histogram Computation. In traditional HoG, an n-bin histogram of gradient in 1D space is counted. Now, we are required to count an n-bin histogram of gradient in 2D space. To simplify the representation, the angle distribution can be seen in a spherical coordinate system. In the spherical coordinate system, the axes are the gradient of the light field in horizontal, vertical and depth directions, i.e. (dx, dy, dz). As a result, there are two angles we should concern, as shown in Fig.2. We can easily know that $\theta \in [0, \pi]$ and $\phi \in (0, 2\pi]$, then divide them into bins. Given a point , it must belong to a partition. Then we can weight their votes into orientation cells. The weight can be described as follows,

$$r = \sqrt{dx^2 + dy^2 + dz^2}$$
$$\theta = arccos\frac{dz}{r}, \phi = arctan\frac{dy}{dx} \quad (4)$$

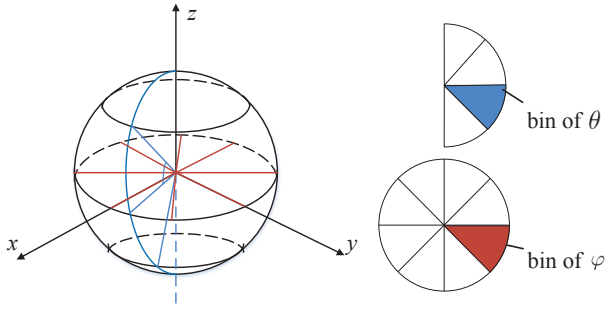
According to the process mentioned before, the histogram h_c can be represented by Eq.5

$$h_c = \sum_{r \in c} r \quad (5)$$

Because the gradient varies over a wide range owing to local variations in illumination and foreground-background contrast, a local contrast normalization turns out to be essential. Here we use L_2 norm to normalize the histogram.



(a) The spherical coordinate system of the LFHoG



(b) An illustration of bins of θ and ϕ

Fig. 2. An illustration of LFHoG representation

Descriptor Synthesis. Similar to traditional HoG in 2D imagesgiven a light field Image I, we divide I into $S*S$ blocks b_i, and then the b_i can be divided into $Q*Q$ cells c_j, besides, the adjacent blocks are overlapping, as shown in Fig.3. So all non-overlapping cells form the set over which the histogram is computed. For each cell, an orientation histogram is computed (Eq.4) and normalized. All histograms are finally concatenated into a feature vector $h=\left(h_{c_1}, h_{c_2}, ..., h_{c_{S^2*Q^2}}\right)$. Our experiments show that the configuration of $Q = 3$ and cell size with 8 pixels can achieve best performance, where the missing rate reaches the minimum.

3. EXPERIMENTAL RESULTS

We have validated the proposed LFHoG descriptor on the dataset that we have taken with Lytro camera. Our dataset contains 92 live faces, 84 pad faces, and 300 negative data without face. Fig.4 shows some examples of light field images. Another thing which must be mentioned is that we acquire the pad face by taking pictures with face shown in the pad with Lytro. Then we process the dataset as gray image to simplify the process.

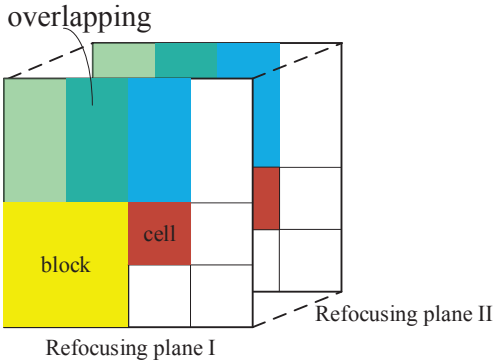


Fig. 3. The block and cell partition in a light field.

(a) (b) (c)

Fig. 4. The training samples in the experiments. (a) is the training samples of live face. (b) is the training samples of pad face. (c) is the training samples of non-face.

In our experiment, the template size is fixed at and our detection window is . We can scale the test image to detect the live face in it. Due to the limitation of running time and storage, we set the scale step as 0.7. Besides, the detection step is set as 4 pixels, that is to say, the adjacent detection windows are at the distance of 4 pixels.

Considering the simplicity and speed, we use linear SVM in libsvm [17] as a baseline classifier throughout the study. Live face detection test was performed on a machine with 6.00GB RAM and Intel Cores i3 CPU in a 64-bits windows system.

To verify the accuracy of the model, we have done a 6-fold cross validations, we partition our training data as six parts which include 16 live face, 14 pad face and 50 scenes without any human face. Then we select the five parts for training and left one parts for testing. The result is shown in Tab.1. The best result on test data is 100%. What's more, the ROC curve our method get on the test cases is shown in Fig.5;

Table 1. The accuracy of live face classification

Cross-validation	1	2	3	4	5	6
Accuracy(%)	100	99.75	99.75	99.75	99.75	99.75

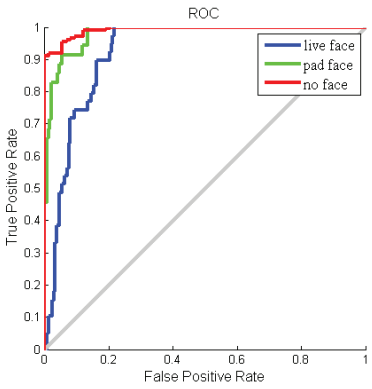
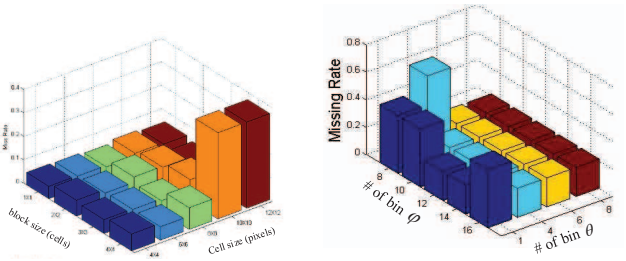


Fig. 5. The ROC curve of our method

There are three parameters to affect the performance of detection, including the block size, cell size and the number of bins. We will discuss their influence respectively.

Fig.6(a) shows the missing rate with different block sizes (cell number) and cell sizes (in pixels). For light field detection, 3*3 cell blocks of 8*8 pixel cells perform best. In other words, when the block size and cell size are too small, the structure information of the whole scene will be lost with more time spent, but the detail will be lost with the sizes are too large, so the appropriate sizes selected is reasonable.

Fig.6(b) shows the missing rate with the changing of bin numbers. In general, the more the bin numbers are, the missing rate will be the lower. With the tradeoff of the time and accuracy, the bin number of ϕ is set as 16 and the bin number of θ is set as 4.



(a) (b)

Fig. 6. The performance with different parameters. (a) is the result of the influence with different block sizes (cell number) and cell sizes (in pixels). (b) is the result of the influence with different bin numbers.

Fig.7 shows some results of LFHoG descriptors. But there are still some cases that our descriptor works not well, as shown in Fig.7(d). As the Fig.7(c) shown, the detection result can be resulted by the gestures of the human and the pad face.

Through our analysis, we can find that the close-up face can be detected very well since the disparity of face is significant, but the live face would be seen as a flat when it is far. So we do an experiment to present the change of accuracy with the depth increasing, as Fig.8 shown. The main reason is the

(a) (b) (c) (d)

Fig. 7. The results of LFHoG descriptors. (a) The result of only the pad face. (b) The result of only the live face. (c) The result of both live face and pad face in one image. (d) The wrong detection cases of face missing.

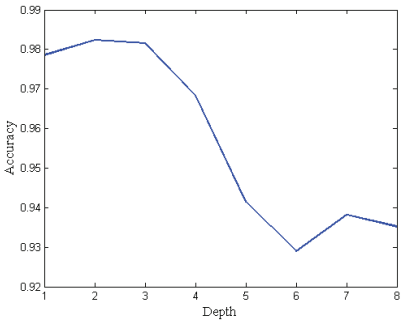


Fig. 8. The change of accuracy with the depth increasing.

limitation of the axial resolution in the Lytro camera, which can influence the precision of refocusing and may generate little difference between the two refocusing focal plane.

4. CONCLUSION AND FUTURE WORK

In the paper, we propose a novel light field descriptor (LFHoG) based on the histogram of three directions (vertical, horizontal, depth). With the joining of the gradient along depth direction, the LFHoG merges the structure information of the rendered image and the distribution of the scene depth together, which is a more comprehensive description of the light field. To verify the effectiveness of the LFHOGG descriptor, we apply it in the live face detection. The experiments are carried out on the dataset we took with light field camera. We also evaluate influences of the parameters and optimize them for live face detection. The accuracy of live face detection using the LFHoG feature can reach 99.75% averagely.

In future work, we will try to apply the descriptor in more application of light field, such as plane detection, action recognition and so on.

5. REFERENCES

[1] S Wanner and B Goldluecke, "Variational light field analysis for disparity estimation and super-resolution.," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 36, no. 3, pp. 606–19, 2014.

[2] T. E. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *IEEE Trans Pattern Anal Mach Intell*, vol. 34, no. 5, pp. 972–986, 2012.

[3] Patrick Hanrahan and Ng Ren, "Digital light field photography," *Dissertation Abstracts International, Volume: 67-05, Section: B, page: 2664.;Adviser: Patrick Hanra*, vol. 115, no. 3, pp. 38–39, 2006.

[4] Changil Kim, Henning Zimmer, Yael Pritch, Alexander Sorkine-Hornung, and Markus Gross, "Scene reconstruction from high spatio-angular resolution light fields," *Acm Transactions on Graphics*, vol. 32, no. 4, pp. 96–96, 2013.

[5] 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*, 2014, pp. 1 – 8.

[6] B. Peixoto, C. Michelassi, and A. Rocha, "Face liveness detection under bad illumination conditions," in *IEEE International Conference on Image Processing*, 2011, pp. 3557–3560.

[7] Xiaoyang Tan, Yi Li, Jun Liu, and Lin Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *European Conference on Computer Vision*, 2010, pp. 504–517.

[8] Jiamin Bai, Tian Tsong Ng, Xinting Gao, and Yun Qing Shi, "Is physics-based liveness detection truly possible with a single image?," in *International Symposium on Circuits and Systems*, 2010, pp. 3425 – 3428.

[9] Gang Pan, Lin Sun, Zhaohui Wu, and Yueming Wang, "Monocular camera-based face liveness detection by combining eyeblink and scene context.," *Telecommunication Systems*, vol. 47, no. 3-4, pp. 215–225, 2011.

[10] O. V. Komogortsev and A. Karpov, "Liveness detection via oculomotor plant characteristics: Attack of mechanical replicas," in *International Conference on Biometrics*, 2013, pp. 1–8.

[11] Prathap Nair and Andrea Cavallaro, "3-d face detection, landmark localization, and registration using a point distribution model," *IEEE Transactions on Multimedia*, vol. 11, no. 4, pp. 611–623, 2009.

[12] S Kim, Y. Ban, and S Lee, "Face liveness detection using a light field camera.," *Sensors*, vol. 14, no. 12, pp. 22471–99, 2014.

[13] R. Raghavendra, Kiran B. Raja, and Christoph Busch, "Presentation attack detection for face recognition using light field camera.," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 1060–75, 2015.

[14] A. Ghasemi and M. Vetterli, "Detecting planar surface using a light-field camera with application to distinguishing real scenes from printed photos," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2014, pp. 4588–4592.

[15] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conference on Computer Vision & Pattern Recognition*, 2005, pp. 886–893.

[16] Marc Levoy, Ren Ng, Andrew Adams, Matthew Footer, and Mark Horowitz, "Light field microscopy," *Acm Transactions on Graphics*, vol. 25, no. 3, pp. 924–934, 2006.

[17] Chih Chung Chang and Chih Jen Lin, "Libsvm: A library for support vector machines," *Acm Transactions on Intelligent Systems & Technology*, vol. 2, no. 3, pp. 389–396, 2011.