DETECTING PLANAR SURFACE USING A LIGHT-FIELD CAMERA WITH APPLICATION TO DISTINGUISHING REAL SCENES FROM PRINTED PHOTOS

Alireza Ghasemi *Martin Vetterli*

AudioVisual Communications Laboratory
School of Computer and Communication Sciences
École Polytechnique Fédérale de Lausanne

ABSTRACT

We propose a novel approach for detecting printed photos from natural scenes using a light-field camera. Our approach exploits the extra information captured by a light-field camera and the multiple views of scene in order to infer a compact feature vector from the variance in the distribution of the depth of the scene. We then use this feature for robust detection of printed photos.

Our algorithm can be used in person-based authentication applications to avoid intruding the system using a facial photo. Our experiments show that the energy of the gradients of points in the epipolar domain is highly discriminative and can be used to distinguish printed photos from original scenes.

Index Terms— Light-Field Imaging; Plenoptic Function; Feature Extraction.

1. INTRODUCTION

Password based authentication has been shown to be vulnerable in many applications. Recently, in many modern digital devices such as smartphones or laptops, there is an authentication option to replace the password-based authentication with personal face detection-based authentication mechanisms[1]. However, most of such systems suffer from vulnerabilities to intrusion by using a printed photo of the authorized face.

The main reason for such a security weakness in person-based authentication systems is the lack of depth information in traditional color cameras. Consider for example the two images shown in Figure 1. One of them has been taken from a natural scene and the other is the result of scanning a printed photograph. One can hardly say which one is the natural scene and which one is the print. This example shows that in the image plane, there are few discriminating features available to distinguish a scanned photo from a natural scene.

Thus, in order to be able to distinguish and reject printed photos, we have to capture more than visual data using the sensing device. Modern depth cameras such as the Microsoft Kinect [2] or thermal cameras [3] have been commercialized in recent years and can be used to capture the information necessary to detect printed photos. However, the main problem with such cameras is that the power consumption and the dimensions of the device prevent it from being embedded into mobile consumer devices such as smartphones or tablets.

Fig. 1: Two Images. One from a natural scene and the other from a printed photo.

Another option for extracting extra information from the scenes is using a light-field camera [4]. Light field cameras have received wide attention in recent years due to the commercialization of consumer oriented products such as the Lytro [5]. Light-field analysis has been applied to various problems in computer vision from face recognition [6] to depth estimation [7]. Moreover, successful efforts have been made to embed the light-field imaging capability into mobile devices [8]. Therefore, it is likely that we will have light field cameras in cellphones and we can exploit the extra information these cameras provide from the surrounding scene in order to improve current computer vision methods.

In this paper we propose a simple and efficient algorithm for planar surface detection using a light field camera which can be used for detecting printed photos from natural scenes. Several methods has been recently proposed in the literature for flat surface detection. In [9], authors propose to detect flat grounds using the disparity map of the scene. Their approach works well for robot stereo vision. However, it is restricted to detecting grounds and can not be generalized to other parts of the scene. Moreover, the algorithm is sensitive to stereo alignment errors.

In [10], a system is proposed for contour detection in digital TVs. However, their approach relies on color information

Thanks to the CTI for funding this project.

and does not exploit the depth of the pixels.

[11] analyzes three methods for defect detection in surfaces. The proposed methods depend on laser scanners and controlled industrial environments in order to work properly.

Our proposed approach can robustly and reliably distinguish flat surfaces from natural, deep scenes. The algorithm can be applied to a consumer light-field camera and does not rely on expensive laser scanners or calibration-required stereo cameras. Moreover, the algorithm is robust ro various deformations in the surface such as tilting. Another benefit of our approach is that it doesn't require modifying existing face databases in authentication systems which can be time consuming and expensive.

An important application for our algorithm could be detecting intruders in face detection-based authentication systems. Since light-field cameras are likely to appear in mobile devices in the near future, our algorithm can be easily incorporated into future personal authentication systems of mobile phones and other hand-held devices.

2. THE PROPOSED APPROACH FOR DETECTING FLAT SURFACES

2.1. The Plenoptic Function and Formation of Light-Fields

A Plenoptic function is a generalization of a two-dimensional image, in which we have as well as spatial coordinates of the image plane, five more dimensions for the time(t), wavelength (λ) and the position of the camera ($\langle V_x, V_y, V_z \rangle$) [12]. This leads to:

$$P = P_7(x, y, V_x, V_y, V_z, t, \lambda). \tag{1}$$

Due to the high-dimensionality of the plenoptic function, we usually capture and work with a certain subset of it. A commonly known 3D restriction of the plenoptic function is the $x - y - V_x$ slice. This is known as the Epipolar Plane Image [13] or the light-field [14]. Light-fields have attracted a lot of attention in recent years [15, 16, 17, 18].

Assuming the pinhole camera model and Lambertian surfaces [19], consider an image sequence, taken by moving the camera V_x units along the horizontal axis for each image (i.e the light-field case). Adding a third dimension for the camera location (set to V_x), the mapping from a scene point $P = (X, Y, Z)^T$ to its projection into each image in the sequence can be described as:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \mapsto \begin{pmatrix} f\frac{X}{Z} - f\frac{V_x}{Z} \\ f\frac{Y}{Z} \\ V_x \end{pmatrix}. \tag{2}$$

This is how a light-field is formed. We infer from (2) two important facts. First, each scene point is mapped to a line (the epipolar line [20]) in its corresponding $x - V_x$ slice (known as the EPI plane [13]). Moreover, the gradient (slope) of an epipolar line is proportional to the depth (Z value) of the scene.

We will use these properties in deriving our approach for detecting printed photos from natural scenes.

2.2. Detecting Flat Surfaces Using the EPI Information

Suppose we capture two images using a light-field camera from the two seces in the Figure 1. In Figure 2, we see two of the epipolar planes (i.e. $x - V_x$ slices) of the images in figure 1. Now the difference becomes visible: We observe that for the printed photo, all epipolar lines in a plane have the same gradient (slope[1]).

This is absolutely predictable if we recall that the gradient of an epipolar line is proportional to the depth of its corresponding point. This is not the case for a natural scene in which every scene point may potentially lie in a different depth layer and thus the gradients of the lines vary dramatically.

However, the gradients of epipolar lines will not remain the same if we tilt the printed photo in front of the camera. Therefore, a single plane can not provide a robust measure for detecting printed photos. Another important property seen in light field images of the printed photos in figure 2 is the invariance of lines' gradient distribution among the EPI planes. This property is also easily verified by noting the depthlessness property of this type of light fields. We use this property to overcome the problem of non-precise line detection in images. We will use this property in our solution.

The approach we propose is based on extracting a feature vector called the energy feature vector v from all epipolar planes of a light field. The energy feature vector measures the energy or variance of the gradient in the dominant lines among all epipolar planes of a light field. The size of the energy vector is the number of top dominant lines in the plane which are considered.

Therefore, the first step of our algorithm is detecting and estimating the orientation of the lines in each epipolar plane. For the line detection phase, we use the exponential Radon transform (ERT) introduced in [21]. The exponential radon transform is a variant of the well-known Radon transform which is redesigned and optimized for EPI plane analysis. The main benefit of ERT is its parameter space which directly involves lines' gradients (i.e. slopes) which correspond to depth of scene points, rather than orientations which do not have direct meaning in the EPI domain.

After applying the line detection transform to all epipolar planes, we compute the amount of change (i.e. the energy) in the gradient of top k dominant lines along all EPI planes. Concatenated together, the energy of gradients form our k-dimensional energy feature vector.

The overall algorithm for planarity based feature extraction in a light field is depicted in Algorithm 1.

[1] In this paper we will use the terms gradient and slope interchangeably

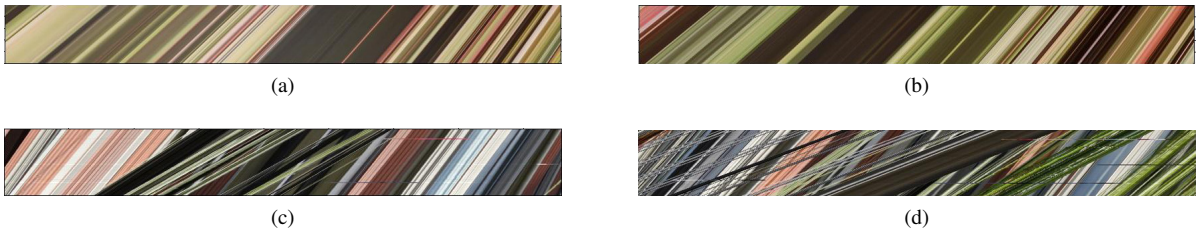

Fig. 2: Four sample EPI planes of the two light fields in figure 1. 2a and 2b has been captured from the printed photo while 2c and 2d are from the natural scene.

Input: The three-dimensional light field volume L.
Output: The number of dominant lines k.
begin
  foreach *EPI plane E in L* **do**
    Construct the parameter space H_E from E using the exponential Radon transform.
    Select the k largest elements in H_E.
    Construct the orientation vector O_E `/* Use the orientation (gradient) of the top k elements. */`
  end
  Construct the orientation matrix O `/* Each row of O is the orientation vector of one of the EPI planes. */`
  return *the energy vector v constructed by computing the energy of each column of O.*
end

Algorithm 1: The Algorithm for Computing the Energy-Based Feature Vector from a Light-Field

There are two main reasons why we use an energy based method on all epipolar planes. First, the error in the line detection algorithm makes it difficult to have all dominant lines in the same direction. Moreover, the photo under consideration may not be completely fronto-parallel. Thus, there may be actual changes of gradient in an epipolar plane.

Figure 3 depicts the gradients of top five dominant lines in all EPI planes of the natural scene in Figure 1. Horizontal axis is the index of the dominant lines and vertical axis shows the variance. We can see the high variation in the gradients.

3. EXPERIMENTAL RESULTS

3.1. The Experimental Setup

For evaluating the accuracy of the proposed system distinguishing natural scenes from printed photos, we used the light field data captured by the Disney research lab Zurich [14]. The Disney research data have been captured using a simple generic setup which allows us to assess the quality of our method regardless of the different camera architectures of current light-field systems . We implemented the energy computation algorithm in the Matlab 2013a environment, on a computer with 8 gigabytes of RAM and a dual-core processor.

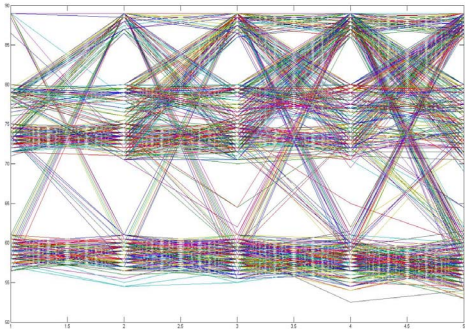

Fig. 3: Gradients of dominant lines in all EPI planes

There are light-field data of five captured scenes in the Disney research dataset: Mansion, Church, Bikes, Couch and statue. We resized all images to have Full HD resolution (1920×1080 pixels). The number of views (2D images) in each light-field varies from 51 to 100 images. For each of the scenes, we created a *copy* light-field by capturing a moving camera image sequence of the printed photo of the middle frame in the *original* light-field. We executed the experiments with different values of k to see the effects.

3.2. Analysis of the Energy Feature Vectors

The ℓ^2-norm of the extracted energy feature vectors for each of the scenes are depicted in Table 1 for $k = 5$ and $k = 10$. We can see that the ℓ^2-norms for copy light-fields are close to zero, close to each other and distinct from those of original light-fields.

Regarding the effect of k, we see in table 1 that increasing the value of k causes increase in the ℓ^2-norm of variations. The main reason for this phenomenon is that the line detection error is higher when detecting parameters (including orientation) of the less dominant lines since they are shorter (composed of less points).

Figure 4 depicts the k-dimensional energy vector corresponding to the original and copy light-fields of the Disney research dataset, for $k = 5$. Here we again see that components in the copy light-field are very close to zero.

Scene	Copy_5	Original_5	Copy_{10}	Original_{10}
Mansion	0.2879	37.87	0.65	67.78
Couch	0.4563	18.94	5.13	22.47
Church	0.3156	48.05	16.44	54.12
Statute	0.3459	33.63	2.27	83.01
Bikes	0.3211	16.12	0.87	29.34

Table 1: ℓ^2-norm of energy feature vectors for the Disney Research light fields.

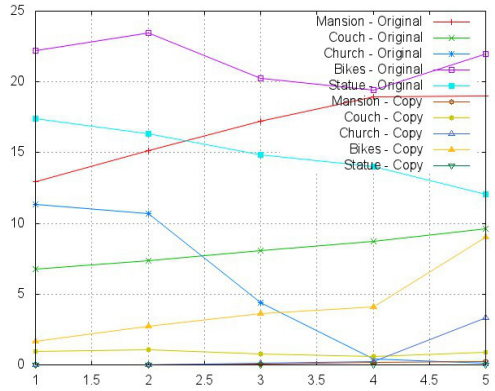

Fig. 4: The 5-dimensional energy feature vector for the Disney light-field dataset

3.3. Further Experiments and Comparison with the State-of-the-art

In order to further validate our algorithm and assess its robustness, we created a dataset of various light-fields, both flat and non-flat (i.e. natural). To capture the dataset, we used the Lytro consumer light-field camera [5]. The dataset consists of 50 light-fields of printed photos and 50 light-fields of natural scenes. We applied various degrees of tilting to the printed light-fields to evaluate the robustness of our algorithm. We set the distance between the camera and the scene from 5 to 25 centimeters which is comparable to the case when one want the camera to do face detection. Although the Lytro has a low angular resolution, it captures enough depth variations for the energy vectors to work, as we will see below..

In the Table 2, we see the performance of our approach on the Lytro dataset as a confusion matrix. To model and test our feature vectors in a classification setting, we used Bayesian classification with leave-one-out cross-validation [22].

We can see in the Table 2 that the energy vectors perform very well in detecting printed photos using a low angular resolution commercial Lytro camera. The results is important since it shows that the algorithm does not need a wide baseline to perform well in practical applications. Moreover, we used a very simple pattern recognition algorithm such that the performance of the feature vectors is not influenced by the machine learning part of our approach. Higher accuracy may be achieved by using more robust recognition algorithms such as Gaussian processes or the SVM [22].

Regarding failure cases, we observed that they mostly occur in situations where the distance of the scene from the camera is far more than the camera baseline and therefore the resulting depth variation is less than what is required by the algorithm to be detected as natural (non-flat).

	C/EV	O/EV	C/HOG	O/HOG
Copy (C)	46	4	28	22
Original (O)	7	43	23	27

Table 2: Comparison of detection accuracy between our approach (EV) and the Histogram of Oriented Gradients (HOG) on the dataset of Lytro images.

The Dataset	$k = 5$	$k = 10$
Original Images HD	3.92 ± 0.14	6.02 ± 0.22
Copy Images HD	5.65 ± 0.19	6.34 ± 0.17
Original Images Full HD	22.76 ± 0.33	21.03 ± 0.12
Copy Images Full HD	23.49 ± 0.45	23.88 ± 0.52

Table 3: The time complexity of energy feature computation. Time is measured in seconds.

We also compared our results with the state-of-the-art widely-used Histogram of Oriented Gradients (HOG) feature descriptor [23]. Our results show superior performance over HOG. This is mostly because our algorithm exploits light-field information whereas HOG doesn't.

Table 3 depicts the time complexity of the feature vector computation for the Disney dataset. The time complexity of the feature extraction depends on the size of the image and the number of sub-views. The results in Table 3 show that the time complexity is not highly dependent on the value of k. The reason for this could be that the most time-consuming part of the algorithm which is the exponential radon Transform, runs independently of the value of k.

4. CONCLUSION AND FUTURE WORK

Light-field cameras are interestingly popular and will be showing in mobile industry in the future. Our proposed approach can be a reliable way of improving the security of user authentication systems exploiting the unique features of these new cameras. Moreover, there are many other application of the proposed method which can improve upon current computer vision algorithms. For examples, detecting billboards or windows in outdoor scenes can be improved by using a light field camera and exploiting the property of same-orientation lines.

The algorithm for computing energy vectors and performing the final detection can be improved in many ways. First, the energy can be replaced by other diversity-sensitive measures such as PCA. The line detection algorithm may also be replaced by a more robust one or by orientation detection algorithms. The average orientation of edges in the each EPI plane can also be used as a measure of how depth distribution changes in each EPI plane.

We may also study the limits of the algorithms, for example in terms of the number of sub-aperture views.

References

[1] Alaa E. Abdel-Hakim and Motaz El-Saban, "Face authentication using graph-based low-rank representation of facial local structures for mobile vision applications," in *ICCV Workshops*. 2011, pp. 40–47, IEEE.

[2] Billy Y. L. Li, Ajmal S. Mian, Wanquan Liu, and Aneesh Krishna, "Using kinect for face recognition under varying poses, expressions, illumination and disguise," in *WACV*. 2013, pp. 186–192, IEEE Computer Society.

[3] Reza Shoja Ghiass, Ognjen Arandjelovic, Hakim Bendada, and Xavier Maldague, "Vesselness features and the inverse compositional aam for robust face recognition using thermal ir," in *AAAI*, Marie desJardins and Michael L. Littman, Eds. 2013, AAAI Press.

[4] Tom E. Bishop and Paolo 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.

[5] Todor Georgiev, Zhan Yu, Andrew Lumsdaine, and Sergio Goma, "Lytro camera technology: theory, algorithms, performance analysis," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2013, pp. 86671J–86671J.

[6] R. Raghavendra, Kiran B. Raja, Bian Yang, and Christoph Busch, "A novel image fusion scheme for robust multiple face recognition with light-field camera," in *FUSION*. 2013, pp. 722–729, IEEE.

[7] Sven Wanner, Christoph N. Straehle, and Bastian Goldluecke, "Globally consistent multi-label assignment on the ray space of 4d light fields," in *CVPR*. 2013, pp. 1011–1018, IEEE.

[8] Kartik Venkataraman, Dan Lelescu, Jacques Duparr, Andrew McMahon, Gabriel Molina, Priyam Chatterjee, Robert Mullis, and Shree Nayar, "Picam: An ultra-thin high performance monolithic camera array," in *ACM SIGGRAPH 2013 Asia*. ACM, 2013.

[9] Jun Zhao, Mark Albert Whitty, and Jayantha Katupitiya, "Detection of non-flat ground surfaces using v-disparity images," in *IROS*. 2009, pp. 4584–4589, IEEE.

[10] Wonseok Ahn and Jae-Seung Kim, "Flat-region detection and false contour removal in the digital tv display," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005, pp. 1338–1341.

[11] Pingbo Tang, Burcu Akinci, and Daniel Huber, "Characterization of three algorithms for detecting surface flatness defects from dense point clouds," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2009, pp. 72390N–72390N.

[12] Edward H. Adelson and John Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 99–106, 1992.

[13] R.C. Bolles, H.H. Baker, and D.H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1, no. 1, pp. 7–55, 1987.

[14] Changil Kim, Henning Zimmer, Yael Pritch, Alexander Sorkine-Hornung, and Markus Gross, "Scene reconstruction from high spatio-angular resolution light fields," *To appear ACM Trans. Graph.(Proc. SIGGRAPH)*, 2013.

[15] Ivana Tosic, Sapna A. Shroff, and Kathrin Berkner, "Dictionary learning for incoherent sampling with application to plenoptic imaging," in *ICASSP*. 2013, pp. 1821–1825, IEEE.

[16] Minh N. Do, Davy Marchand-Maillet, and Martin Vetterli, "On the bandwidth of the plenoptic function," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 708–717, 2012.

[17] Yosuke Bando, Henry Holtzman, and Ramesh Raskar, "Near-invariant blur for depth and 2d motion via time-varying light field analysis," *ACM Trans. Graph.*, vol. 32, no. 2, pp. 13, 2013.

[18] Kshitij Marwah, Gordon Wetzstein, Yosuke Bando, and Ramesh Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 46, 2013.

[19] J. Berent and P.L. Dragotti, "Segmentation of epipolar-plane image volumes with occlusion and disocclusion competition," in *Multimedia Signal Processing, 2006 IEEE 8th Workshop on*. IEEE, 2006, pp. 182–185.

[20] J. Berent and P.L. Dragotti, "Plenoptic manifolds," *Signal Processing Magazine, IEEE*, vol. 24, no. 6, pp. 34–44, 2007.

[21] A. Ghasemi and M. Vetterli, "Scale-invariant representation of light-field images for object recognition and tracking," *Electronic Imaging 2014*, February 2014.

[22] C.M. Bishop and SpringerLink (Service en ligne), *Pattern recognition and machine learning*, vol. 4, springer New York, 2006.

[23] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 886–893.