USER-ORIENTED STEREO VIDEO REFOCUSING BY COMPUTATIONAL CINEMATOGRAPHIC MODEL

Wenjing Geng, Dapeng Du, Tongwei Ren and Gangshan Wu

State Key Laboratory for Novel Software Technology
Nanjing University, Nanjing 210023, China
jenngeng@gmail.com, dudp.nju@gmail.com, {rentw, gswu}@nju.edu.cn

ABSTRACT

Refocusing, as a most popular photographic technique, is widely welcomed in both photography and cinematography. Currently, most refocusing effects in movies are implemented manually using professional devices or computer graphics methods. Few solutions are designed for the videos captured by common users. In this paper, we propose a user-oriented method to facilitate the video refocusing for daily life, with an extra requirement for only stereo cameras which are quite popular today. Given a stereo video, one can select the desired focusing part in a single frame, and the user selected intention will be tracked along with the timeline. Then the depth-of-field (DoF) based blurring effect can be generated by scene depth estimation based on the stereo techniques and the proposed cinematographic models in computational photography. The performance of the refocused videos compares favourably with the ones generated by digital single lens reflex (DSLR). To evaluate the proposed method, we conduct the experiments on several challenging stereo video datasets. The quantitative and subjective evaluations both show that the proposed method can achieve attractive and highly aesthetic performance with few user interactions.

Index Terms— Photography, video refocusing, stereo camera, computational cinematographic model

1. INTRODUCTION

Knowing how to make the parts of your image you want sharp and the parts you want to be out of focus, is a great artistic tool to create great images, as well as videos [1, 2]. The emergence on stereoscopic equipment breeds large amount of stereo media in recent years. With the increasing binocular handheld devices, such as Fujifilm 3D camera, Huawei honor and the upcoming dual-lens smartphones, 3D videos have quietly stepped into people's daily life, which is no longer a kind of media living in the cinemas or 3D TVs. Precisely because of the widely use of stereo devices, it is a fair chance of achieving computational photography without the help of professional camera devices, such as DSLR. It is noted that, refocusing, as a most popular photographic and cinematographic technique, is pursued by many shutterbugs and is also applied in most movies to create romantic and aesthetic atmosphere [3]. However, depending on the professional equipment and specialized knowledge would make video refocusing extravagant for ordinary users. Inspired by building a bridge between the refocusing and the binocular video captured by widely used 3D equipment, and making 3D videos much more vivid and interesting, a user-oriented stereo video refocusing framework is proposed in this paper aiming at generating attractive refocusing results specified by user intention and re-edited by the presented cinematographic model. A user can also perform multiple interactions to select different focus planes for each frame to create the rack focusing effect, as shown in Fig. 1.

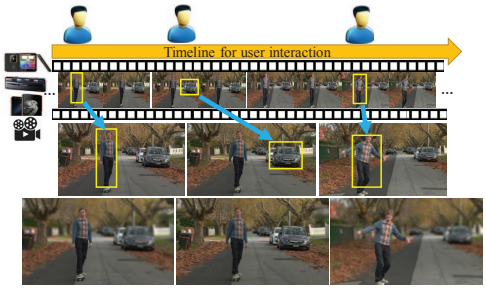


Fig. 1. Interactive video refocusing facilitated to ordinary users. From top to down are the stereoscopic videos, user multiple interactions along with the video timeline, and the refocused results generated by the proposed method.

Some previous works involving video refocusing [4, 5] have achieved satisfactory refocusing results by designing customized equipment for acquiring depth more accurately. The extra overhead of setting hardware limits the popularization of these techniques and generates bottlenecks. Besides, the categories of the experimental videos are less than abundance and detached from the shoot scenes of life. Some other methods [6, 7] calculated depth by utilizing the parameters of the cameras or the smartphones, and the movements during the acquisition of the video. These approaches suffer from the disadvantages of stipulating restricted conditions. Furthermore, the refocused continuity should also be considered in the scheme of video refocusing,

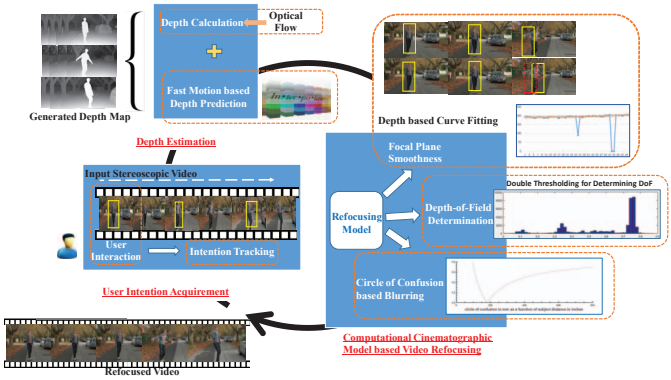


Fig. 2. The overview of the proposed framework.

or the flickers between successive frames would highly degrade the results. Motivated by the above challenges, we propose a user-oriented video refocusing framework by considering both consistency and artistic style. First, a user interaction is provided to get the sharp part specified by the user. And the subsequent intentions are tracked by online learning detection [8] for its high efficiency. Then the depth maps are calculated on the stereo videos. We present a motion based depth prediction method to preserve the continuous changes of depth as much as possible and improve the computing efficiency meanwhile. To create a comparable refocusing as captured by professional equipment, a computational cinematographic model is designed to offer constraints when generating the final results, including depth based focus trajectory smoothness, DoF range determination and circle of confusion (CoC) based blurring. The overview of the proposed framework is depicted in Fig. 2.

To evaluate the performance, a relatively perfect public stereoscopic video dataset [9] with affluent scene categories but without ground truth of depth is treated as user inputs. And the experimental results show that the proposed approach performs well on such a challenging dataset. In summary, the main contributions are summarized as follows.

- An effective user-oriented framework for video refocusing, providing a user interface that doesn't require professional knowledge to use, is proposed in this paper which takes both feasibility and effectiveness into consideration;
- A motion-based depth prediction is presented to preserve the continuity among consecutive frames while speeding up the depth calculation;
- A computational cinematographic model is constructed considering the rules of filming by professional cameras, including focus smoothness, DoF determination and CoC based blurring.

2. RELATED WORK

In this section, some representative works about video refocusing are firstly outlined. And then some depth estimation methods are reviewed in a brief way.

Video refocusing is an artistic performance in photography while an interesting research topic in computational photography. Moreno-Nogue et al. [4] proposed a system for refocusing images and videos of dynamic scenes using a single-view depth estimation method. The depth was estimated by a single camera and the defocus of a sparse set of dots projected onto the scene. Therefore, a hardware system is essential to get accurate depth maps. Kuthirummal et al. [5] presented an imaging system enabling one to control DoF by designing an ingenious camera. The accurate refocusing results are guaranteed by the arrangements of lens. Shroff et al. [6] presented the variable DoF imaging as example, but this paper mainly aimed at improving the traditional depth from defocus (DFD) to dynamic scenes. Therefore, it has some defects inheriting from DFD, such as the need for a lens with large enough aperture to cause DoF, which is not applicable to the ordinary users. Yu et al. [7] introduced a synthetic aperture effect for images, but this paper focused on proposing a novel depth estimation method in some specific cases. Tambe et al. [10] presented an implementation of an LF video camera, and did video refocusing based on the proposed camera. In a word, the existing video refocusing methods mainly rely on accurate depth estimation under some specific conditions. And the experiments are not mainly conducted on the videos captured in real life.

Depth estimation, as a basic processing in 3D scene understanding, has become an active research topic these years. Different from structured light [11], time-of-fight [12] and laser scanning [13], we estimate the depth by binocular vision, which is a low cost method close to human eyes. Stereo matching [14, 15] is widely used in depth calculation for stereo images. However, the usual requirements for appropriate disparity ranges limit to getting the optimized results and applying stereo matching frame by frame would lead to discontinuity of depth even in some consecutive sequences. Few depth estimation for stereo videos exist. Wedel et al. [16] computed depth between neighboring two frames, which still contains computing redundancy. Valgaerts et al. [17] utilized four-frame configuration which has not solved the consistency very well. To preserve

temporal consistency in depth maps, Hung et al. [18] proposed an optimized method with a bunch of constraints and optimizations. The high computing efficiency of these methods makes it emergent to find a feasible and effective solution served for the application proposed in this paper.

3. USER-ORIENTED VIDEO REFOCUSING

Inspired by building a bridge between stereoscopic media captured by widely used stereo devices and refocusing effect being famous in photography, we propose a user-oriented video refocusing framework shown in Fig. 2.

3.1. User intention tracking

The proposed framework aims at imitating a famous artistic effects, refocusing, controlled by cinematographers. Generally, the cameraman can skillfully design the distribution of both clear and unclear parts in one shot. However, for an ordinary user, it is hard to apply much photography knowledge when recording a video. And the aesthetic standards for determining the clear scope of regions vary from person to person. Therefore, most videos captured by ordinary users are all focused. In spite of these issues, users still favor refocusing techniques very much.

In order to fully satisfy the user intention, we offer a user interaction process aiming at achieving customized user intention of where to be clear. Firstly, the user is required to draw a rect R as the initialized intention on the initial video frame of one view where should be kept clear in the subsequent sequences. Then a track learning detection framework proposed in [8] is introduced for its efficiency and good performance on fulfilling the realtime and long-term tracking task. These tracked regions R are the sets of the parts of the focus planes for the video, and denoted as Eq. (1):

$$R = \{R(c_1, 1), R(c_2, 2), ..., R(c_i, i), ...R(c_n, n)\} \quad (1)$$

where $R(c_i, i)$ represents the focus plane focused by point c_i for frame i, and $R(c_1, 1)$ is initialized by user interaction.

It is noteworthy that the tracking method proposed in [8] has the deficiency of inaccurate object localization and tracking loss, which would mislead the user intention tracking in the subsequent frames. Therefore, we propose a focal plane smoothness based user intention tracking, which is further described in Sec. 3.3.1.

3.2. Video depth estimation

For keeping the consistency of depth and improving the computing efficiency as much as possible, we propose an optical flow based depth prediction method. First, a fast optical flow algorithm [19] is performed between sequential frames i and $i+1$ to get a series of motion fileds $v = [v_x, v_y]^T$ for each frame. Second, the feature descriptors of histogram of oriented optical flow (HOOF) are extracted to represent the motion fields of each frame i, represented as $h_i = [h_{i;1}, h_{i;2}, ..., h_{i;bin}]^T$. Then the prediction label of the next frame is determined by the following decision function:

$$p(i) = \begin{cases} 1, & S(h_i, h_{i-1}) < \tau \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $S(h_i, h_{i-1})$ measures the similarity between successive HOOF and when it is below τ, the depth map of frame $i+1$ can be predicted by motion-based interpolation. Let D_i be the depth map of frame i, the predicted value of the next depth map can be calculated by Eq. (3).

$$D_{i+1}(x, y) = D_i(x - v_x, y - v_y) \quad (3)$$

If $p(i)$ of the frame i equals 0, the depth of this frame is calculated by the optical flow estimation method proposed in [20] which performs very well on the depth boundary and the textureless region with a little high computing cost. And the depth maps are refined by a hole filling method based on the previous frames.

3.3. Computational cinematographic model

Cinematography is the science or art of motion photography, recording dynamic pictures by means of professional equipment with an image sensor and the light-sensitive material [21]. There are many aesthetic standards observed by an excellent cinematographer. However, we aim at simplifying the procedures for the ordinary users to touch refocusing. By leveraging three main factors having influence on the professional effects, the continuity of the focus, the range of the DoF and the pattern of blur, we propose a computational cinematographic model.

3.3.1. Focal plane smoothness

The focal is a point located on the optical axis of a rotationally symmetric. A good refocused video should have an organized change of focus, i.e. the trajectory of focus in depth space among the neighboring frames should present disciplinary changing. It is determined by the usage of camera. In our model, we posit that the regions with the mode of the depth in $R(c_i, i)$ is the exact focus plane specified by the user. The depth based focus plane F_i can be denoted as Eq. (4).

$$F_i = \{f_m | f_m \in R(c_i, i), D(f_m) = Mode(R_i, D)\} \quad (4)$$

where f_m is the m_{th} pixel in the tracked rect $R(c_i, i)$ whose depth value is the mode of all the depth values in $R(c_i, i)$ of frame i. By user intention tracking, we found that the track learning detection method could not localize focus plane very well at every frame. The depth value of the focus in each rect would vary a lot or get lost in some frames. The unexpected weak coupling in user intention tracking highly degrades the artistic style during the overall video flow. In order to smooth the changing trajectory among focus planes, we use a surface fitting method based on the depth of focus point to readjust

the focus planes. Finally we got a refined focus plane set $\{F_1, F_2, ..., F_n\}$ for each frame in the time domain.

3.3.2. Depth-of-field determination

A basic definition of DoF is the zone of acceptable sharpness within a photo that will appear in focus. Hence, DoF is a clear scope determined by aperture, distance from the subject to the camera, and focal length of the lens on the camera in photography. Therefore, automatically calculating DoF is the first step for achieving refocusing. As we provide a user interaction as the first process, the DoF can be estimated from the region where user select. Here we propose a depth mode aided double thresholding algorithm based on Otsu, which is an effective automatic thresholding method widely used in image segmentation.

Assume that the depth scale of user selected region is $[0, S]$, meanwhile each bin has n_{d_i} pixel number with the depth value d_i. The Otsu algorithm is improved to obtain thresholds of d_{min} and d_{max} which represent the start-stop values of DoF. Firstly, t_1 is calculated to divide the depth histogram into $T_a = [0, t_1]$ and $T_b = [t_1 + 1, S]$ by maximizing the inter-class difference between T_a and T_b. The optimized calculation process can be denoted as:

$$t_1 = \arg\max \sum w_a w_b (\overline{d_a} - \overline{d_b}) \tag{5}$$

where w_a and w_b respectively equal n_a/n and n_b/n. n is the pixel number in the user selected rect. $\overline{d_a}$ and $\overline{d_b}$ are the mean depth value of T_a and T_b. Secondly, T_a or T_b are divided to T_{a1}, T_{a2} or T_{b1}, T_{b2} based on the comparison with the mode of depth $d_m(R)$ restricted by rect R. The second threshold can be elucidated as Eq. (6) or Eq. (7).

$$t_2 = \arg\max \sum w_{a1} w_{a2} (\overline{d_{a1}} - \overline{d_{a2}}), \quad d_m(R) < t1 \tag{6}$$

$$t_2 = \arg\max \sum w_{b1} w_{b2} (\overline{d_{b1}} - \overline{d_{b2}}), \quad d_m(R) > t1 \tag{7}$$

where w_{a1} and w_{a2} respectively equal n_{a1}/n_a and n_{a2}/n_a, w_{b1} and w_{b2} respectively equal n_{b1}/n_b and n_{b2}/n_b. $\overline{d_{a1}}$ and $\overline{d_{a2}}$ are the mean depth value of T_{a1} and T_{a2}. $\overline{d_{b1}}$ and $\overline{d_{b2}}$ are the mean depth value of T_{b1} and T_{b2}. Then the start-stop range of DoF can be denoted as:

$$DoF = \begin{cases} [0, t_2 + \xi] & 0 < d_m(R) < t_2 < t_1 < S \\ [t_2 + \varepsilon, t_1 + \xi] & 0 < t_2 < d_m(R) < t_1 < S \\ [t_1 + \varepsilon, t_2 + \xi] & 0 < t_1 < d_m(R) + \xi < t_2 < S \\ [t_2 + \varepsilon, S] & 0 < t_1 < t_2 < d_m(R) < S \end{cases} \tag{8}$$

where the start-stop value of $[.]$ represents the start-stop value of DoF. ε and ξ are the slack parameters to adjust the refocusing results. For the DoF is not always equally distributed in front and back of the focus point. It is usually about one third in front and two thirds behind the focal point. But it becomes more equal when the focal length increases. By inducing two slack parameters, the way of refocusing infinitely is close to the DSLR principle which is essential in the proposed model.

Table 1. Description of the stereo video dataset

Source	Num.	Resolution	Shot types
Uni. Nantes [23]	12	1920*1080	tracking, no move
RMIT [22]	24	1920*1080	pan, no move, dollying tilt, pull back, tracking
Others	11	1920*1080	pan, dollying, no move

3.3.3. Circle of confusion based blurring

Circle of confusion is a concept in optics. It is caused by a cone of light rays from a lens not coming to perfect focus when imaging a point source. From the view of optics, it is an optical spot, while from the view of imagery, it is a blur spot. In Sec. 3.3.2, the DoF is calculated by the depth mode aided double thresholding algorithm. However, DoF is just a range being clear and is not enough to determine the degree of blurring. In order to build a relationship between depth and blurring, we adopt CoC based blurring to create a real feeling. Let C represent the radius of the circle of confusion, the relationship of C and each depth value $D(x, y)$ can be denoted as Eq. (9):

$$C = A \frac{|D(x, y) - f|}{2D(x, y)} \tag{9}$$

where A is the aperture diameter in the camera and can be estimated by fixing a value with a fixed range of DoF for C. f is the distance of the focused object plane that can be got the same way as A. In optics, these variables are usually calculated in millimeter, we just do with them in the pixel scale. By utilizing Eq. (9), the blurring level for each depth is calculated being used as the blurring circle for each pixel.

4. EXPERIMENTS

4.1. Experimental settings

We evaluate our approach on an integrated dataset [9] collected from multi sources, the RMIT3DV database [22], the IVC stereoscopic video database of University of Nantes [23], the Elephant Dream project and the others. The dataset includes 47 stereo videos without depth ground truth but with the computed depth maps. We use 32 bins for HOOF. And the other parameters of our method are set as $\{\tau, \varepsilon, \xi, S\} = \{0.05, 0, 0, 255\}$ throughout the experiments. All the experiments are conducted on a machine with a 3.4GHz Intel i7-4770 CPU and 16GB memory. We simply classify the dataset into 6 common types, pan shot, dollying shot, tilt shot, pull back shot , track up shot and no camera movement shot, based on the shot type classification in cinematography. As shown in Table 1, the dataset we used covers all the basic shot types in cinematography.

4.2. Results and discussions

A thorough comparison is conducted by considering the influence of depth, focus, and the pattern of blurring. And the

user satisfaction is investigated to evaluate the artistic results.

Depth continuity The providers of the public dataset offered the calculated depth maps, shown in Fig. 3 (b). Fig. 3 (c) shows the depth maps calculated frame by frame. It could be seen that the depth maps offered by the dataset and calculated by [20] frame by frame obviously show the discontinuity even among the consecutive frames and the depth values on the edges performed not very well. Some comparable details are enlarged in the last row of Fig. 3. The final depth maps generated by the proposed motion-based interpolation method can eliminate this kind of discontinuity shown in Fig. 3 (f). We also demonstrate each processing results in Fig. 3 (d)-(f), including depth generated by motion interpolation, after hole filling and by color image guided filtering. It is worth noted that the depth maps shown in Fig. 3 (f) are all generated by the interpolation.

Focus smoothness Due to the analysis of temporal imaging characteristics and the probable moving trajectory of the camera, the frame content customized by users, i.e. the fictitious focus plane, should not have the abrupt changes among neighboring frames unless users perform multiple interactions. In brief, the focus trajectory should present a gradual and organized changing in the depth space. However, after the initial user intention tracking, shown in the second line of Fig. 4, the focus plane would not cover the user intention very well, and even encounter missing matching in several frames, which would highly degrade the refocusing results illustrated in the third line of Fig. 4. To fix this issue, we employ a surface fitting to refine the focus plane in each frame, demonstrated in the fourth line of Fig. 4. The improved refocusing results are exhibited in the last line of Fig. 4. We also draw the depth-time curves of the videos with and without focus smoothness in Fig. 4. In this way, the user intention can be well unscrambled in each frame.

Pattern of blurring Gaussian blur is a famous method for blurring an image in image processing. An intuitive idea for refocusing is that making the part specified by the user clear while making the other part obscure by Gaussian blur. Although it can simply generate the refocusing-like effect, depicted in Fig. 5 (a), the blurring degree maintains the same in each depth level, which violates the light diffraction principle in optics. In order to create a DSLR-like blurring, we employ a circle of confusion based pattern for it accords with the defocus mechanism occurred in camera. Fig. 5 (b) shows the CoC based blurring, with marked objects by rectangles from near to far. It could be seen that the CoC based blurring varies from the distance.

User satisfaction To explore the novelty and popularity of the proposed application, we invite nine volunteers (three females and six males) who know nothing about the photography skills to participate in user feedback study based on a scoring system for evaluating six different methods of rendering the refocusing videos. The methods are offered depth based video refocusing by

(a) Gaussian blur based refocusing (b) Circle of confusion based refocusing

Fig. 5. The comparison of pattern of blurring, (a) obscured by Gaussian filter and (b) by circle of confusion. Distance varies from the blue to red with blue rectangle selected by the user.

(a) Scores for each method given by 9 subjects (b) Scores for each method averaged by 9 subjects

Fig. 6. The quantitative statistics for user satisfaction.

Gaussian filter (offer.D.+Gau.), by CoC blurring both without focus smoothness and with smoothness (offer.D.+CoC., offer.D.+CoC.+Smooth.), the proposed calculated depth based refocusing by Gaussian filter (our.D.+Gau.), by CoC blurring both without focus smoothness and with smoothness (our.D.+CoC., our.D.+CoC.+Smooth.). Each result for every video scores by the five-grade marking system and the evaluation of each method is built on a hundred-mark system by offering 20 videos for every subject. Then each method obtains a hundred-mark based score by calculating sums of scores of testing videos as shown in Fig. 6 (a). The final score of a method is averaged by the sum of scores given by each volunteer, as shown in Fig. 6 (b). Obviously, the proposed method got the highest scores for attractive effects on imitating DSLR-like refocusing. We find that the key factor affecting the refocusing results lies in the accuracy of depth map, and the ordinary users are not able to recognize the different effect blurred by Gaussian filter or CoC processing for human eye sensitivity to image blurring is limited, unless he knows something about DSLR principles.

5. CONCLUSION

We have presented a novel refocusing method for stereo videos. The DoF region can be customized by the user interaction. Based on the refocusing knowledge in photography, we build a cinematographic model to guide the refocusing approaching to the DSLR results. The experiments and user satisfactions elucidate the effectiveness of the proposed method.

Acknowledgements. This work is supported by the National Science Foundation of China under Grant No.61321491 and No.61202320. Research Project of Excellent State Key Laboratory (No.61223003), and Collaborative Innovation Center of Novel Software Technology and Industrialization.

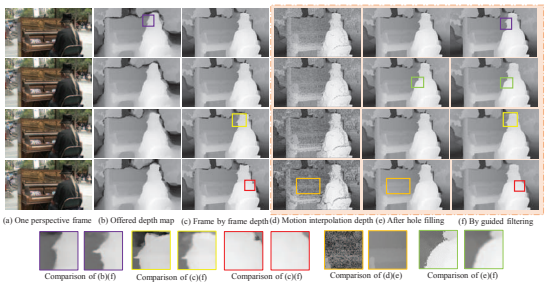
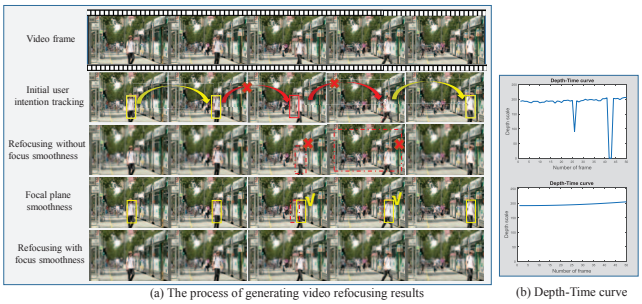


Fig. 3. The comparison of the calculated depth maps by [9], [20] and the proposed method, separately shown in (b), (c), (f). (a) is a perspective for several consecutive frames of one video. (d)-(f) are the processing details of the proposed method. The last row shows some enlarged comparisons as example.



(a) The process of generating video refocusing results (b) Depth-Time curve

Fig. 4. The comparison between video refocusing without and with focal smoothness, (a) illustrates the initial tracking results of user intention and the improved effects by utilizing focal plane smoothness, red rectangle means missing match while yellow is the correct tracking results, and (b) is the depth-time curve before and after focal plane smoothness.

6. REFERENCES

[1] Blain Brown, *Cinematography: theory and practice: image making for cinematographers and directors*, Taylor & Francis, 2013.

[2] Alex Garcia, *Depth of Field: Tips on Photojournalism and Creativity*, Agate Digital, 2014.

[3] David Jones, *Mastering Camera Aperture: Digital Photography Tips and Tricks for Beginners on How to Control Depth of Field*, 2012.

[4] Francesc Moreno-Noguer, Peter N Belhumeur, and Shree K Nayar, "Active refocusing of images and videos," in *TOG*. ACM, 2007.

[5] Sujit Kuthirummal, Hajime Nagahara, Changyin Zhou, and Shree K Nayar, "Flexible depth of field photography," *TPAMI*, 2011.

[6] Nitesh Shroff, Ashok Veeraraghavan, Yuichi Taguchi, Oncel Tuzel, Amit Agrawal, and Rama Chellappa, "Variable focus video: Reconstructing depth and video for dynamic scenes," in *ICCP*. IEEE, 2012.

[7] Fisher Yu and David Gallup, "3d reconstruction from accidental motion," in *CVPR*. IEEE, 2014.

[8] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas, "Tracking-learning-detection," *TPAMI*, 2012.

[9] Yuming Fang, Junle Wang, Jing Li, Romuald Pépion, and Patrick Le Callet, "An eye tracking database for stereoscopic video," in *QoMEX*. IEEE, 2014.

[10] Salil Tambe, Ashok Veeraraghavan, and Ankit Agrawal, "Towards motion aware light field video for dynamic scenes," in *ICCV*. IEEE, 2013.

[11] Matthew O'Toole, John Mather, and Kiriakos N Kutulakos, "3d shape and indirect appearance by structured light transport," in *CVPR*. IEEE, 2014.

[12] Hyunjung Shim and Seungkyu Lee, "Hybrid exposure for depth imaging of a time-of-flight depth sensor," *Optics express*, 2014.

[13] Kourosh Khoshelham and Sander Oude Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, 2012.

[14] Jure Žbontar and Yann LeCun, "Stereo matching by training a convolutional neural network to compare image patches," in *CVPR*. IEEE, 2015.

[15] Andreas Geiger, Martin Roser, and Raquel Urtasun, "Efficient large-scale stereo matching," in *ACCV*. Springer, 2011.

[16] Andreas Wedel, Clemens Rabe, Tobi Vaudrey, Thomas Brox, Uwe Franke, and Daniel Cremers, *Efficient dense scene flow from sparse or dense stereo data*, Springer, 2008.

[17] Levi Valgaerts, Andrés Bruhn, Henning Zimmer, Joachim Weickert, Carsten Stoll, and Christian Theobalt, "Joint estimation of motion, structure and geometry from stereo sequences," in *ECCV*. Springer, 2010.

[18] Chun Ho Hung, Li Xu, and Jiaya Jia, "Consistent binocular depth and scene flow with chained temporal profiles," *IJCV*, 2013.

[19] Ce Liu, *Beyond pixels: exploring new representations and applications for motion analysis*, Ph.D. thesis, Citeseer, 2009.

[20] Deqing Sun, Stefan Roth, and Michael J Black, "A quantitative analysis of current practices in optical flow estimation and the principles behind them," *IJCV*, 2014.

[21] Douglas Arthur Spencer, *The Focal dictionary of photographic technologies*, Prentice-Hall, 1973.

[22] Eva Cheng, P Burton, Jonathan Burton, Alex Joseski, and I Burnett, "Rmit3dv: Pre-announcement of a creative commons uncompressed hd 3d video database," in *QoMEX*. IEEE, 2012.

[23] Matthieu Urvoy, Marcus Barkowsky, Romain Cousseau, Yao Koudota, Vincent Ricorde, Patrick Le Callet, Jesus Gutierrez, and Narciso Garcia, "Nama3ds1-cospad1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3d stereoscopic sequences," in *QoMEX*. IEEE, 2012.