Detecting Momentary Shadows from Visible and Thermal Image Pair

Kazuya Fujita
Department of Artificial Intelligence
Kyushu Institute of Technology
Iizuka, Japan
k_fujita@pluto.ai.kyutech.ac.jp

Ryo Matsuoka
Faculty of Environmental Engineering
The University of Kitakyushu
Kitakyushu, Japan
r-matsuoka@kitakyu-u.ac.jp

Takahiro Okabe
Department of Artificial Intelligence
Kyushu Institute of Technology
Iizuka, Japan
okabe@ai.kyutech.ac.jp

***Abstract*—Outdoor shadows can be classified into two categories: continuous shadows caused by static objects and momentary shadows caused by moving objects. Since the momentary shadows such as shadows due to a photographer are annoying and do not exist in the original scene, they should be detected and removed for improving image quality. In this paper, we propose a method for detecting momentary shadows from a visible and thermal image pair. The key idea of our proposed method is that the continuous shadows have lower temperature than non-shadow areas, while the momentary shadows have almost the same temperature as the non-shadow areas. Therefore, our method combines the shadow areas detected by using an RGB image and the higher-temperature areas detected by using a thermal image, and then detects the areas of momentary shadows via image segmentation. Through a number of experiments using real visible and thermal image pairs, we show that the combination of visible and thermal images are effective for detecting momentary shadows, and that our method works well for momentary shadows with varying duration time.**

***Index Terms*—Shadow Detection, Thermal Imaging, Image Segmentation**

I. Introduction

Outdoor shadows can be classified into continuous shadows and momentary shadows on the basis of their temporal properties. The former ones are caused by static objects such as buildings and trees, and then they are almost static and change gradually according to the position of the sun. On the other hand, the latter ones are caused by moving objects such as people and cars, and then they are dynamic and change according to the objects as well as the sun.

For example, when we take a picture, the momentary shadows caused by a photographer often cast in a scene of interest. Such momentary shadows are more common when we take a picture outdoors under follow light. Since the momentary shadows are annoying and do not exist in the original scene, they should be detected and removed for improving image quality. In this study, we propose a method for detecting momentary shadows in images taken under sunlight.

Detecting (and removing) shadows is one of the most important research topics in image processing and computer vision, and has been studied for a long time. Shadow detection and removal are useful for illumination estimation [10], [14], [16], [17], shape reconstruction [8], camera calibration [7], and preprocessing for computer vision tasks such as object detection and tracking [2], [12]. Not only physics-based approach [5], [6], [20] but also learning-based approach [4], [9], [13], [18] are proposed recently. Unfortunately, however, continuous shadows and momentary shadows have the same brightness and color, when they cast on the same surface. Therefore, it is difficult, even impossible to distinguish those shadows from ordinary RGB images.

In this paper, we propose a method for detecting momentary shadows from a visible and thermal image pair. The key idea of our proposed method is that the continuous shadows caused by static objects have lower temperature than non-shadow areas, while the momentary shadows caused by moving objects have almost the same temperature as the non-shadow areas. Specifically, our method combines the shadow areas detected by using an RGB image and the higher-temperature areas detected by using a thermal image, and then detects the areas of momentary shadows via image segmentation.

To confirm the effectiveness of our proposed method, we conducted a number of experiments using real visible and thermal image pairs taken outdoors, and evaluated the performance of our method both qualitatively and quantitatively. Furthermore, we study the performance of our method under varying duration time of momentary shadows; the temperature of momentary shadows falls as the duration time increases, and then it becomes difficult to distinguish momentary shadows from continuous shadows.

The main contribution of this study is threefold. First, as far as we know, this is the first method for detecting momentary shadows from one-shot visible and thermal image pair without using temporal information. Our study is a novel application of thermal imaging, which becomes more common recently. Second, we experimentally show that the combination of visible and thermal images are effective for detecting momentary shadows. The experimental results also show that our method works well for momentary shadows with varying duration time. Third, our dataset of visible and thermal image pairs will be made publicly available after publication.

II. Related Work

A. Shadow detection and removal

Conventionally, shadow detection and removal are tackled with physics-based approach. For example, Finlayson *et*

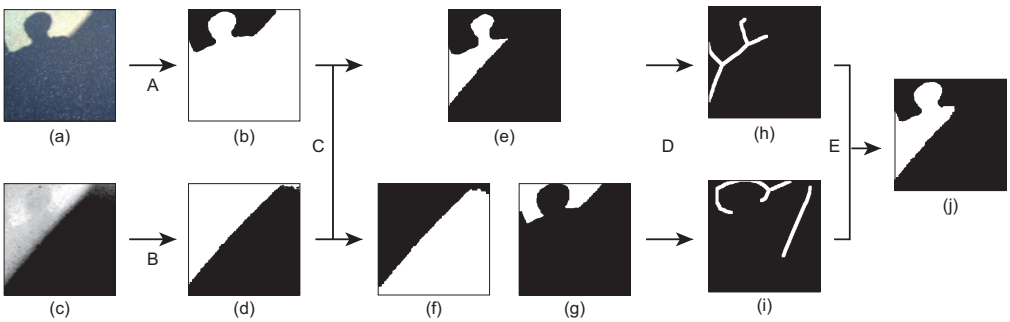

Fig. 1. The flowchart of our proposed method: (a) an input visible image, (b) a shadow map, (c) an input thermal image, (d) a temperature map, (e) foreground candidates, (f)(g) background candidates, (h) initial markings for foreground areas, (i) initial markings for background areas, and (j) a momentary shadow map.

TABLE I
THE KEY IDEA OF OUR PROPOSED METHOD. THE SHADOW AREAS DETECTED BY USING A VISIBLE IMAGE AND THE HIGHER-TEMPERATURE AREAS DETECTED BY USING A THERMAL IMAGE ARE COMBINED.

		thermal	
		higher temperature	lower temperature
visible	shadow	foreground (momentary shadows)	background (continuous shadows)
	non-shadow	background	background

al. [5], [6] derive the invariant direction in a 2D chromaticity feature space, and Yu *et al.* [20] make use of a linear structure in the RGB color space. Recently, learning-based approach achieves better performance especially under uncontrolled conditions. For example, Khan *et al.* [9] use deep CNNs, and Nguyen *et al.* [13], Wang *et al.* [18], and Ding *et al.* [4] propose GAN-based methods: conditional GAN (cGAN), stacked conditional GAN (ST-CGAN), and attentive recurrent GAN (ARGAN) respectively.

Unfortunately, however, it is difficult, even impossible to classify momentary shadows and continuous shadows by using the above methods. This is because they use ordinary color images as input, and momentary shadows and continuous shadows have the same brightness and color when they cast on the same surface.

Our proposed method is similar to Xiao *et al.* [19] in the spirit that a novel modality different from an RGB image is incorporated into shadow detection (and removal). Their method makes use of depth cues and can handle both hard and soft shadows within the same framework. In contrast to the above method, we make use of thermal cues and achieve the classification between momentary shadows and continuous shadows from one-shot visible and thermal image pair without using temporal information.

B. Thermal imaging

A thermal camera can capture the infrared (IR) radiation from an object of interest, while an ordinary color camera captures the reflected light from the object surface illuminated by a light source. Because the SPD (spectral power distribution) of the radiation depends on the temperature of the object, we can measure the spatial distribution of temperature by using thermal cameras. Therefore, thermal images are important for classifying objects with different temperatures.

For example, Davis *et al.* [3] combine visible and thermal images for background subtraction and achieve robust person detection. Bulanon *et al.* [1] combine visible and thermal images for robustly detecting fruits. Thermal imaging is useful also for face detection and recognition in low light conditions. Zhang *et al.* [21] propose GAN-based method for thermal to visible face image conversion.

In this study, we exploit thermal imaging for shadow detection. In particular, we show that thermal cues are effective for classifying momentary shadows and continuous shadows without using temporal information. Our study is a novel application of thermal imaging, which becomes more common as consumer thermal cameras become widespread recently.

III. PROPOSED METHOD

The continuous shadows caused by static objects have lower temperature than non-shadow areas, while the momentary shadows caused by moving objects have almost the same temperature as the non-shadow areas. Therefore, as shown in TABLE I, our proposed method detects momentary shadows from a visible and thermal image pair; it combines the shadow areas detected by using the visible image and the higher-temperature areas detected by using the thermal image, and then detects the areas of momentary shadows via image segmentation.

Specifically, as shown in Fig. 1, our proposed method (A) labels the visible image pixels, (B) labels the thermal image pixels, (C) combines those labels from the visible and thermal images, (D) marks foreground and background areas, and finally (E) segments foreground and background areas via graph cut based image segmentation. The rest of this section describes each step of our method in more detail.

A. Labeling visible image pixels

First, as shown in Fig. 1 A, we label each pixel of (a) an input visible image as shadow or non-shadow, and obtain (b) a shadow map. We can make use of existing techniques for the

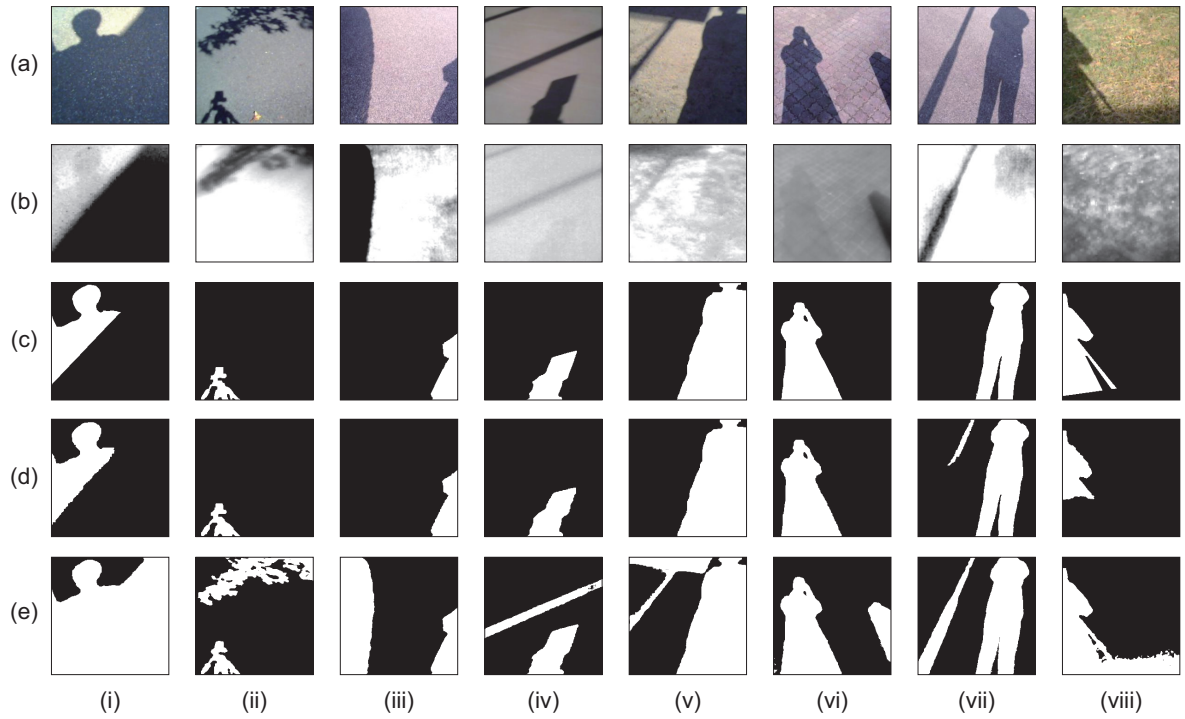

Fig. 2. The qualitative results of our proposed method on 8 scenes: (a) a visible image, (b) a thermal image, (c) the ground truth of momentary shadows, (d) the result of our method, and (e) the result of ST-CGAN [18] for each scene. The background surfaces on which shadows cast are (i)(ii)(iii)(vii) asphalt, (iv) vinyl, (v) sand, (vi) brick, and (viii) grass.

TABLE II
THE QUANTITATIVE RESULTS OF OUR PROPOSED METHOD AND ST-CGAN: THE PRECISION AND RECALL OF THE DETECTED MOMENTARY SHADOWS.

scene		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
ours	precision	0.967	0.932	0.990	0.970	0.970	0.994	0.941	0.985
	recall	0.970	0.986	0.996	0.997	0.986	0.996	0.994	0.672
ST-CGAN [18]	precision	0.242	0.185	0.233	0.534	0.796	0.803	0.734	0.580
	recall	1.000	0.989	0.999	0.997	1.000	1.000	1.000	0.996

labeling of visible image pixels. In our current implementation, we used one of the state-of-the-art methods: ST-CGAN [18]. The white and black pixels in (b) the shadow map stands for shadow and non-shadow pixels respectively.

B. Labeling thermal image pixels

Second, as shown in Fig. 1 B, we label each pixel of (c) an input thermal image as higher temperature or lower temperature, and obtain (d) a temperature map. We assume that a scene of interest has low thermal conductivity, and the temperature at a surface point gradually changes when the momentary shadow due to a moving object casts on the point. Since the shadows caused by the sun is classified into umbra and penumbra, we classify the thermal image pixels into three classes according to the pixel values (temperatures) by thresholding [15]. The white pixels in (d) the temperature map stand for higher-temperature ones (the first and second classes with higher temperatures), and the black pixels stand for lower-temperature ones (the third class with lower temperatures) respectively.

C. Combining labels from visible and thermal images

Third, as shown in Fig. 1 C, we combine the above labels independently estimated from the visible and thermal images, and obtain (e) a foreground map, (f) a background map, and (g) another background map. The foreground pixels, *i.e.* the white pixels in (e), are shadowed and have higher temperature, and therefore they are the candidates for momentary shadows. On the other hand, the first background pixels, *i.e.* the white pixels in (f), are shadowed and have lower temperature, and therefore they are the candidates for continuous shadows. The second background pixels, *i.e.* the white pixels in (g), are considered as non-shadow.

D. Marking foreground and background areas

Forth, as shown in Fig. 1 D, we automatically mark (h) foreground and (i) background areas in order to provide them as the initial conditions of graph cut based image segmentation. Because the labels near the boundaries between black and white areas in (e), (f), and (g) often have low confidence, we

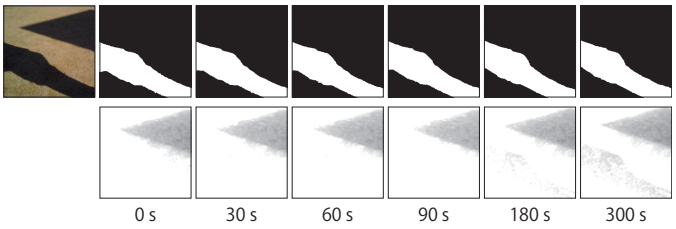

Fig. 3. The results on sand under varying duration time; the visible image and the detected momentary shadows (top) and the corresponding thermal images (bottom).

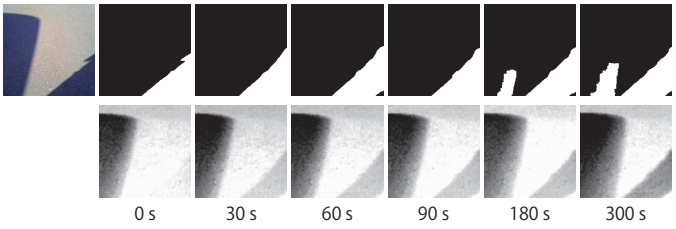

Fig. 4. The results on asphalt under varying duration time; the visible image and the detected momentary shadows (top) and the corresponding thermal images (bottom).

obtain the initial markings of the foreground and background areas by shrinking and thinning those areas.

E. Segmenting foreground and background areas

Finally, as shown in Fig. 1 E, we segment the foreground and background areas from the visible and thermal image pair and the initial markings via graph cut based image segmentation, and obtain (j) a momentary shadow map. We can make use of a variant of existing techniques for graph cut based image segmentation. In our current implementation, we used a variant of Lazy Snapping [11]. Specifically, we use a four-channel image as input; three are from a visible image with RGB channels and one is from a thermal image with a single channel (temperature). The white pixels in (j) the momentary shadow map stand for momentary shadows, and the black pixels stand for continuous shadows and non-shadow areas respectively.

IV. Experiments

To confirm the effectiveness of our proposed method, we conducted a number of experiments using real visible and thermal image pairs taken outdoors. We used PI200 from Optris and C5 from FLIR, which can capture a pair of visible and thermal images at the same time. Because the viewpoints of a thermal camera and a visible camera in PI200 are slightly different, we aligned the captured thermal and visible images by using the homography matrix estimated in advance. The numbers of pixels of thermal images are lower than those of color images (1/16 and 1/4 for PI200 and C5 respectively), and then the boundaries between higher- and lower-temperature areas are blurred to some extent. We show the qualitative and quantitative results on 10 scenes in total, for which we manually labeled the ground truth of momentary shadow areas.

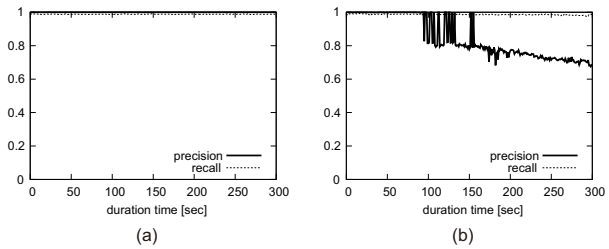

Fig. 5. The precision and recall of the momentary shadows detected on (a) sand and (b) asphalt.

A. Evaluation and comparison

Fig. 2 shows the qualitative results of our method on 8 scenes: (a) a visible image, (b) a thermal image, (c) the ground truth of momentary shadows, and (d) the result of our method for each scene. The background surfaces on which shadows cast are (i)(ii)(iii)(vii) asphalt, (iv) vinyl, (v) sand, (vi) brick, and (viii) grass. Those results qualitatively show the effectiveness of our method; we can see that the momentary shadow areas in our result, *i.e.* the white pixels in (d) are almost the same as those in the ground truth, *i.e.* the white pixels in (c) for the scenes from (i) to (vi). TABLE II shows the quantitative results of our method: the precision and recall of the momentary shadows detected by using our method. Those results quantitatively show the effectiveness of our method; we can see that both the precision and recall are high for the scenes from (i) to (vi).

The failure cases of our proposed method are shown in Fig. 2 (vii) and (viii). In the first case (vii), we can see that a part of the continuous shadows due to a thin object, a pole in this case, are misclassified as momentary shadows. This is because the temperature of such shadows, *i.e.* penumbra are higher than that of umbra as shown in the thermal image. In the second case (viii), we can see that a part of the momentary shadows due to a photographer and a tripod are misclassified as continuous shadows. We can see that the thermal image does not necessarily work well for classifying momentary shadows and continuous shadows in this case, because the leaves of grass cast shadows each other and the temperature of non-shadow areas is not high.

In Fig. 2 and TABLE II, we show the results of ST-CGAN [18], one of the state-of-the-art shadow detection methods, for comparison. Since the purpose of ST-CGAN is shadow detection (and removal), it detects shadows, that include both momentary shadows and continuous shadows, from a visible image as shown in Fig. 2 (e). TABLE II shows that ST-CGAN works well as shadow detector, and achieves high recall. Note that ST-CGAN has higher recall than our proposed method in theory, because the momentary shadows are a subset of shadows detected by using ST-CGAN. On the other hand, the precision of ST-CGAN is small, because it does not distinguish momentary shadows and continuous shadows.

B. Detection under varying duration time

The temperature of momentary shadows goes down as the duration time increases, and then it becomes difficult to distinguish momentary shadows from continuous shadows. Accordingly, we study the performance of our method under varying duration time of momentary shadows. Specifically, we captured the sequence of visible and thermal image pairs for 5 min with the fps of 1 sec, and studied the precision and recall of the momentary shadows detected from each pair by using our method. We tested two scenes whose background surfaces are sand and asphalt respectively.

First, Fig. 3 shows the detected momentary shadows (the white pixels at the top row) on sand under varying duration time. We can see that the momentary shadows are accurately detected when the duration time increases up to 5 min, even though the temperature of the momentary shadows slightly decreases as the thermal images (the bottom row) show. Fig. 5 (a) shows the precision and recall of the detected momentary shadows. We can also see that our method achieves both high precision and high recall when the duration time increases up to 5 min.

Second, Fig. 4 shows the detected momentary shadows on asphalt under varying duration time. We can see that the momentary shadows are accurately detected when the duration time increases up to 90 sec, but a part of continuous shadows is misclassified as momentary shadows for longer duration time. This is because the temperature of momentary shadows on asphalt rapidly falls, and it becomes difficult to distinguish momentary shadows from continuous shadows on the basis of temperature. Fig. 5 (b) also shows that the precision of the detected momentary shadows decreases when the duration time is longer than 90 sec.

Those results show that our proposed method works well under varying duration time of momentary shadows, *e.g.* 90 sec in the above examples. On the other hand, the performance of our method depends on background surfaces, more specifically on the thermal conductivities of background surfaces; asphalt has higher thermal conductivity than sand.

V. Conclusion and Future Work

In this paper, we proposed a method for detecting momentary shadows from one-shot visible and thermal image pair without using temporal information. Our proposed method makes use of the fact that the continuous shadows caused by static objects have lower temperature than non-shadow areas, while the momentary shadows caused by moving objects have almost the same temperature as the non-shadow areas. Specifically, our method combines the shadow areas detected by using the visible image and the higher-temperature areas detected by using the thermal image, and then detects the areas of momentary shadows via image segmentation. Through a number of experiments using real visible and thermal image pairs, we show that the combination of visible and thermal images are effective for detecting momentary shadows, and that our method works well for momentary shadows with varying duration time. Our future study includes the extension to complex background surfaces; we can take the difference in thermal conductivity into consideration via background material segmentation. The extension to artificial light sources such as LEDs and fluorescent lamps with negligible heat radiation is another future direction of this study.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers JP20H00612 and JP17H01766.

References

[1] D. M. Bulanona, T. F. Burksa, and V.Alchanatis, “Image fusion of visible and thermal images for fruit detection”, Biosystems Engineering, 103(1), pp.12–22, 2009.

[2] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, “Improving shadow suppression in moving object detection with HSV color information”, In Proc. ITSC2001, pp.334–339, 2001.

[3] J. W. Davis and V. Sharma, “Background-subtraction using contour-based fusion of thermal and visible imagery”, CVIU, 106(2-3), pp.162–182, 2007.

[4] B. Ding, C. Long, L. Zhang, and C. Xiao, “ARGAN: attentive recurrent generative adversarial network for shadow detection and removal”, In Proc. ICCV2019, pp.10212–10221, 2019.

[5] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, “On the removal of shadows from images”, IEEE TPAMI, 28(1), pp.59–68, 2006.

[6] G. D. Finlayson, M. S. Drew, and C. Lu, “Entropy minimization for shadow removal”, IJCV, 85, pp.35–57, 2009.

[7] I. N. Junejo and H. Foroosh, “Estimating geo-temporal location of stationary cameras using shadow trajectories”, In Proc. ECCV2008 (LNCS 5302), pp.318–331, 2008.

[8] K. Karsch, V. C. Hedau, D. A. Forsyth. and D. W. Hoiem, “Rendering synthetic objects into legacy photographs”, ACM TOG, 30(6), pp.1–12, 2011.

[9] S. H. Khan, M. Bennamoun, F. Sohel, and R. Togneri, “Automatic shadow detection and removal from a single image”, IEEE TPAMI, 38(3), pp.431–446, 2016.

[10] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan, “Estimating natural illumination from a single outdoor image”, In Proc. ICCV2009, pp.183–190, 2009.

[11] Y. Li, J. Sun, C.-K. Tang, H.-Y. Shum, “Lazy snapping”, ACM TOG, 23(3), pp.303–308, 2004.

[12] I. Mikic, P. C. Cosman, G.T. Kogut, and M. M. Trivedi, “Moving shadow and object detection in traffic scenes”, In Proc. ICPR2000, pp.I-321–324, 2000.

[13] V. Nguyen, T. F. Yago Vicente, M. Zhao, M. Hoai, and D. Samaras, “Shadow detection with conditional generative adversarial networks”, In Proc. ICCV2017, pp.4520–4528, 2017.

[14] T. Okabe, I. Sato, and Y. Sato, “Spherical harmonics vs. Haar wavelets: basis for recovering illumination from cast shadows”, In Proc. CVPR2004. pp. I–50–57, 2004.

[15] N. Otsu, “A threshold selection method from gray-level histograms”, IEEE TSMC, 9(1), pp.62–66, 1979.

[16] A. Panagopoulos, C. Wang, D. Samaras, and N. Paragios, “Simultaneous cast shadows, illumination and geometry inference using hypergraphs”, IEEE TPAMI, 35(2), pp.437–449, 2013.

[17] I. Sato, Y. Sato, and K. Ikeuchi, “Acquiring a radiance distribution to superimpose virtual objects onto a real scene”, IEEE TVCG, 5(1), pp.1–12, 1999.

[18] J. Wang, X. Li, and J. Yang, “Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal”, In Proc. CVPR2018, pp.1788–1797, 2018.

[19] Y. Xiao, E. Tsougenis, and C.-K. Tang, “Shadow removal from single RGB-D images”, In Proc. CVPR2014, pp.3011–3018, 2014.

[20] X. Yu, G. Li, Z. Ying, and X. Guo “A new shadow removal method using color-lines”, In Proc. CAIP2017 (LNCS 10425), pp.307–319, 2017.

[21] T. Zhang, A. Wiliem, S. Yang, and B. Lovell, “TV-GAN: generative adversarial network based thermal to visible face recognition”, In Proc. ICB2018, pp.174–181, 2018.