様　式　C－19、F－19－1、Z－19　(共通)

科学研究費助成事業　　研究成果報告書

令和　元 年　 6 月 21 日現在

機関番号：13901
研究種目：基盤研究(C)（一般）
研究期間：2016～2018
課題番号：16K06349

研究課題名（和文）異種性能カメラの相互補完に基づく超高機能多視点映像システムの研究

研究課題名（英文）High spec multi-camera system from heterogeneous camera array

研究代表者
寺谷　メヘルダド（TERATANI, MEHRDAD）

名古屋大学・工学研究科・特任准教授

研究者番号：70554830

交付決定額（研究期間全体）：（直接経費）　　3,600,000円

研究成果の概要（和文）：このプロジェクトの成果は以下のリストの通りです。 (1) 異種カメラを用いた新しい撮影システムの開発。 このマルチカメラシステムにより、研究目的で異種カメラのあらゆる可能な組み合わせをシミュレートすることができます。 (2) マルチスピーチマルチカメラからハイスペックマルチカメラへの変換のためのいくつかのアルゴリズムの開発 - 多項式フィッティングおよび変位推定。 (3) 提案されたアルゴリズムのシミュレーションおよび検証。

研究成果の学術的意義や社会的意義
By this research, several commercial users can share the content that they captured from an event in different angles, by their smart phone and in return they can be given a high-spec multi-view data from several angle and with the high quality in term of resolution and frame-rate.

研究成果の概要（英文）：The achievements of this project are itemized in the list below. (1) Developing a new capturing system using heterogeneous cameras. By this multi-camera system, we can simulate any possible combination of heterogeneous cameras for research purposes. (2) Development of several algorithms - polynomial fitting and displacement estimation, for conversion from multi-spce multi-camera to high spec multi-camera (3) Simulation and verification of the proposed algorithm. (4) Publish and present in Journal, conference and MPEG standardization.

研究分野：Image processing

キーワード：heterogeneous cameras　conversion　polynomial fitting　high spec multi-camera　Simulation

Super multiview video (SMV) contents has been recently become a research platform for experts in 3D imaging. It consists of several hundreds of viewpoint video of a 3D scene. The cost for producing such content is extremely expensive.

(1) This research considers that any type of camera that commercial users have in their hand cab be used as a capturing system of the 3D scene. For example, all people who are capturing the scene of a soccer game in a stadium with their smart phones that have different spec (frame rate, resolution, etc.) are actually capturing super multiview content of that soccer game.

(2) This research aim considers that there will be a solution that with the content of video captured by the acquisition system explains above, we can create a super multiview content that will have the highest spec among those multiview videos.

y@E

(1) This proposal proposes a new capturing system consists of a set of heterogeneous camera array (i.e. camera array consists of cameras with different specifications) that is cost effective. For this research a set of heterogeneous cameras is made for research and proof of concept.
(1a) Also, a different kind of camera (Plenoptic 2.0) has also additionally investigated in this research in order to understand the possibility of having a mix combination of plenoptic and conventional camera array (seed for my future research).

(2) This proposal investigates solutions to reconstruct high specification multiview video from multi-spec multiview video, captured by an array of heterogeneous cameras. This technology can also be used for creation of SMV contents and its compression.
(2a) As for the plenoptic camera, I have investigated the technologies to convert its contents to multiview video (additional research aim during the process of the original research proposal t seed for my future research proposals).

yOQ

(1) A camera setup with seven types of camera is mounted as shown in Figure 1. There are cameras with four different specifications as shown in Table 1. All cameras are connected to one PC as a control unit and data recording module. The cameras are synchronized with a external hardware that linked to all cameras as shown in Figure 1.

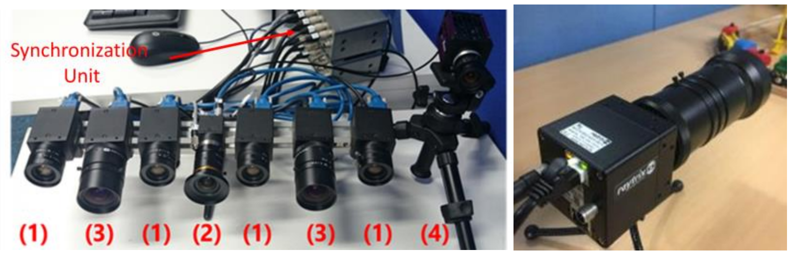

Figure 1. (left) Developed heterogeneous camera array; (right) plenoptic camera type 2.0 (Camera (Raytrix R5-C- GigE-F2.4)

(1a) Additionally, a plenoptic camera with a control unit (PC) is set up for investigation of the capability of such camera for 3D reconstruction and integration to super multiview capturing systems.

Table 1. Specification of each camera in the heterogenous camera array.

Camera	(1)	(2)	(3)	(4)
Resolution	Low	High	Medium	Low
Frame Rate	Medium	Low	High	Medium
# Bits / pixel	Medium	Medium	Medium	High

(2) The proposed algorithm to synthesize a high specification multi-camera system from heterogenous camera array is developed. Figure 2 shows the total algorithm of the proposed method. In order to maximize both resolution and frame rate of cameras a spatial-temporal interpolation is proposed.

In order to perform such interpolation, we need to estimate the displacements (or corresponding points) in each pair of cameras, or multi-camera. Using such information, we can interpolate, any novel view in time or space. Due to differences in views in term of sampling density and field of view, block based matching patch match approaches are not a straight forward solution for estimating a dense displacement. In this research, we propose a scale invariant polynomial expansion method that can estimate a dense displacement between two views, independent of differences in their resolutions, field of views, and their RSWLFDODULVHVbFRQLJXDWLRQ Before any further process, we perform histogram equalization to match the color consistency among cameras.

To overcome these differences in the specification of the camera, we perform a transformation so that all cameras are treated in a same domain independent of their resolutions and frame rates. We propose changing of the coordinate sampled pixels at each view to a common coordinate. In other word, we propose to project pixels at both views to camera coordinate, given the intrinsic parameters of the cameras, as shown in Figure 4.

Now, the projected samples are observed in a common coordinate while the sampling densities are still different. To compare corresponding regions, we approximate coefficients of a quadratic polynomial at each pixel for both views. CompaULVRQ RI WKH SRORQPLDO coefficients is used for estimation of dense displacement. However, when the correspondences between two views have large spatial displacement, such modeling using polynomial expansion cannot perform well, since the method in is originally developed for estimation of displacement in video, where the displacement is mainly small. In order to overcome this problem, we additionally propose multi-resolution analysis for detection of dense correspondence between views.

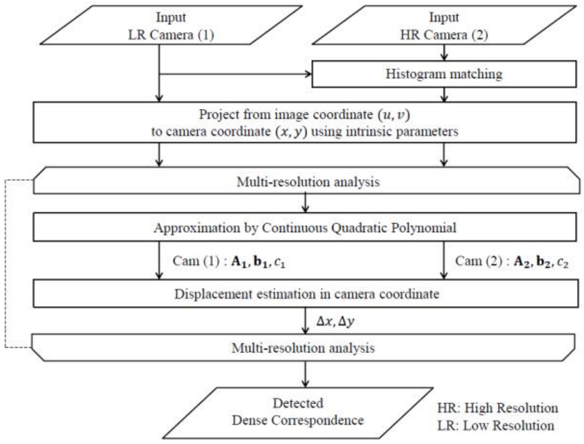

Figure 2. flowchart of the proposed method for spatial-temporal interpolation, given two cameras with different specifications.

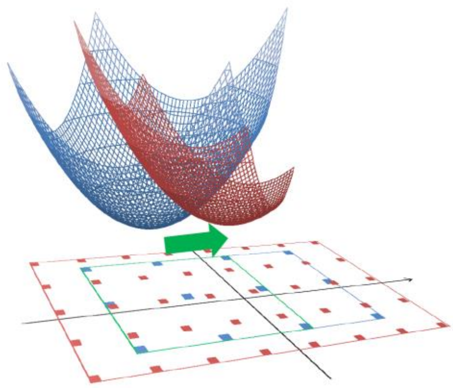

Figure 3. Two images transferred from image to camera coordinate, and approximated quadratic polynomials for the neighborhood. Red dots illustrate the camera image with higher resolution and field of view. Blue dots illustrate the camera image with lower resolution and field of view. Green block illustrates the neighborhood or

(2a) A captured data by plenoptic camera has lenslet format as shown in Figure 6 (more in the homepage -TBA).

In the focused plenoptic camera, each micro-lens observes virtual objects images on a virtual imaging plane. This indicates that each micro-lens acts like a micro camera, and thus, each micro-lens includes spatial and angular information both. Therefore, we should extract a patch not a pixel to render multi-view images. To estimate suitable patch size for every micro-lens, several methods have been investigated. A method proposed depth based method to estimate patch

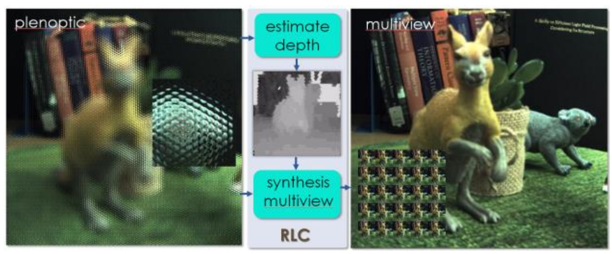

Figure 4. Multiview image from plenoptic camera. RLC is the Reference Lenslet content Convertor; the tool that has been proposed for plenoptic camera

size for every micro-lens. Another method proposed a Laplacian based method to estimate patch size. Based on the later work, we proposed a method using all viewpoint images. In this new method, we applied Laplacian or LoG (Laplacian of Gaussian) instead, after demosaicing the grayscale image. After estimating suitable patch maps, we connect different patch by up-sampling to render lens type, and then use a hybrid method integrate these lens type to obtain multi-view images. Figure 4 briefly explains how the correct patches are selected from different micorlenses in plenoptic camera and created the multiview images.

\TN

(1) Several Data captured by both heterogeneous camera array and the plenoptic camera that are shown in this section and previous section.

(2) I use the estimated displacement from Camera (1) to Camera (2), to synthesize the Camera (2) view, from Camera (1), as shown in Figure 5. Since the synthesized image at Camera (2) is by Camera (1), it does not have all pixels, and it is hard to see the difference in results, we also provide objective measures for this experiment. We used Peak Signal to Noise Ratio (PSNR) for evaluating the accuracy of estimated displacement, objectively. The PSNRs are calculated for available pixels in synthesized Camera (2) in comparison with original Camera (2). In order to make a better understanding of the proposed method, we have performed simulation with modified data of a camera array multiview images that are converted to heterogenous camera array so that subjective and objective evaluation can be performed. In this simulation, the setup is shown in Figure 6 for temporal interpolation. The results of experiments for temporal interpolation are shown in Figure 7.

Figure 5. (left) original image; (left) synthesized image, PSNR = 27.4 dB.

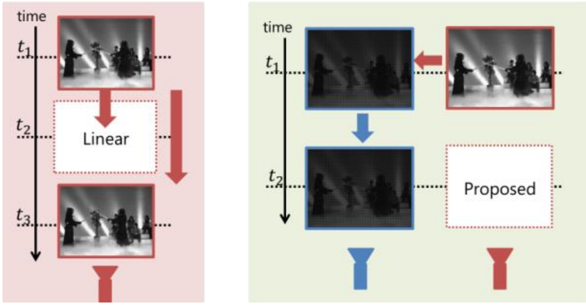

Figure 6. (left) Liner temporal interpolation, (right) proposed temporal interpolation.

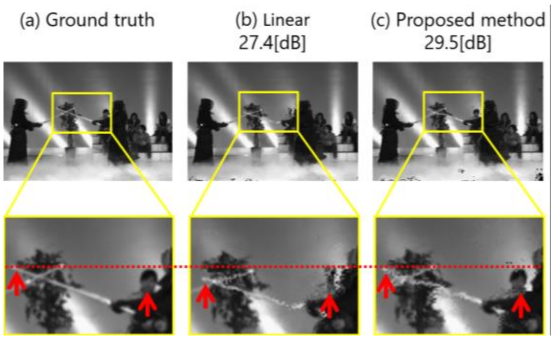

Figure 7. Comparison of the results of temporal interpolation.

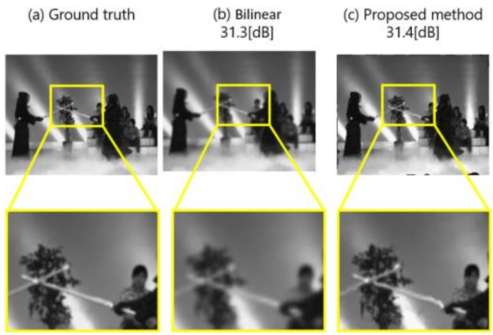

Figure 8. Comparison of the results of spatial interpolation.

In Figure 8, we show the results of experiments for the proposed spatial interpolation with bilinear interpolation for the same purpose. In Figure 9 the results of Spatiotemporal Interpolation is shown that shows that we having heterogenous data we can synthesize all viewpoints with the highest specification.

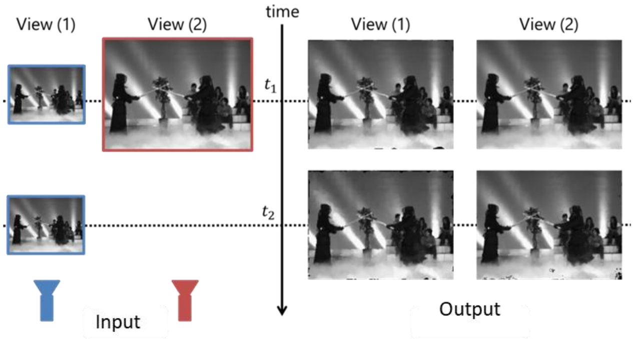

Figure 9. Results of spatial-temporal interpolation using the proposed method.

(2a) The conversion from plenoptic camera to multiview is done by after estimation of the patch map of microlenses. Having the patch map and the plenoptic image, the multiview video can be created, as shown in Figure 10.

Figure 10. Multiview images converted from plenoptic camera

uE₺ u

/ᾹⓈ 2 G¥

K. Shibata, Mehrdad Panahpour Tehrani, 7DNDKDVKL DQ 7)XMLL n(VWLPDWLRQ RI 'HQVH Displacement by Scale Invariant Polynomial Expansion of Heterogeneous Multi-YLHZ ,PDJHV| IEICE Transactions ED, pp. 2048-2051, E100t D, No.9, doi.org/10.1587/transinf.2016PCL0008, Sep 2017.

Mehrdad Panahpour Tehrani, Tomoyuki Tezuka, K. Takahashi, and T. Fujii n)UHH-YLHZSRLQW ,PDJH 6\QWKHVLV 8VLQJ 6XSHUSL[HO 6HJPHQWDWLRQ| $36,3$ 7UDQVDFWLRQV RQ Signal and Information Processing, 6. doi:10.1017/ATSIP. May 2017.

$ᾹⓈ G¥

Mehrdad Teratani 6K\XQLWD :HQ\XH 2Z\QQ .HLWD 7DNDKDVKL 7RVKLDNL)XMLL n' Imaging System Using Multi-)RFXV 3OHQRSWLF &DPHUD DQG 7HQVRU 'LVSOD\| 3URF ,QWHUQDWLRQDO Conference on 3D Immersion (IC3D2018), Brussels, Belgium

Mehrdad Teratani 6K\XQLWD :HQ\XH 2Z\QQ HLWD 7RVKLDNL)XMLL n 03(* -I Visual] Conversion Tool t 5DWULF\DWD WR 0XOWLYLHZ 9LGHR| ,62,(& -7& 6& Macao, SRA China, October 2018.

Tetta Maeda, Ryo Suenaga, Kazuyoshi Suzuki, Mehrdad Panahpour Tehrani, Keita 7DNDKDVKL 7RVKLDNL)XMLL n)UHH 9LHZSRLQW 9LGHR IRU 6SRUWV (YHQWV 8VLQJ 0XOWL-resolution Visual Hull and Micro-IDFHW %LOOERDUGLQJ| 3URF ,QWO :RUNVKRS RQ 6PDUW ,QIR-Media System in Asia (SISA 2016), Ayutthaya, Thailand. (2016.9) (Best Paper Award)

@ ‡ Ø Ȩ Ø ő ‡ _ , nµtUy ţyu2¥W| , j@ìi@ (PCSJ2016 IMPS2016), (2016.11)

0 G¥

ʈ8

0 G¥

E

“

lœ

(ç

y. .

0 G¥

E

“

lœ

ç

y. .

y §

http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data/Raytrix/Tunnel_Train_2/DenseLightFiled_Tunnel_Train_2.htm

http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data/Raytrix/RLC0.2.zip

(1)

v=Ì

Ł

lœ 8 =¥

(2)

â\ v \ z| \ y Ù s Ü vUOq ð dyrdyj| \ y ð N y v

oOqz| y Cgu v . pzrzuy N v d at Ü z| z v > d }