Vniversitat de València

Doctoral Programme in Physics

Doctoral Thesis

Fusion of computed point clouds and integral-imaging concepts for full-parallax 3D display

Seokmin Hong

Supervisors: Dr. Manuel Martínez Corral
Dr. Genaro Saavedra Tortosa
December 2019

D. Manuel MARTÍNEZ CORRAL y D. Genaro SAAVEDRA TORTOSA, Catedráticos de Óptica, adscritos al Departamento de Óptica y Optometría y Ciencias de la Visión de la Universitat de València,

CERTIFICAN

que la presente memoria, *"Fusion of computed point clouds and integral-imaging concepts for full-parallax 3D display"*, resume el trabajo de investigación realizado, bajo su dirección, por D. Seok-min HONG y constituye su Tesis para optar al Grado de Doctor en Física por la Universitat de València.

Y para que conste, en cumplimiento de la legislación vigente, firman el presente certificado en Valencia, a 30 de septiembre de dos mil diecinueve.

Dr. Manuel Martínez Corral   Dr. Genaro Saavedra Tortosa

To all those who believe in me

“And, when you want something, all the universe conspires in helping you to achieve it.”

– Paulo Coelho, *‘The Alchemist’*.

“**R = VD** (**R**ealization = **V**ivid **D**ream)”

– Ji-Sung Lee, *‘Dreaming Attic’*.

Acknowledgements

I affirm that this thesis which entirely inherents my efforts, traces, and outcomes during my doctoral study would become a precious stepping stone for my journey of the future. However, this is by no means that not an achievement and a result of accomplishment by myself, but is come from many of those who believes and encourages me sincerely. I could say that I am tremendously lucky person that I have been able to reach here by virtue of the numerous people's help, the support, the encouragement, the opportunities, and as well as the great teachings. Moreover, lots of beneficial influences and great inspirations during last years gave me a motivation to immerse myself entirely in studies and researches, and eventually, those things have been able to give me a chance for the accomplishment of the doctorate course. Therefore, I sincerely would like to express my deep gratitude for all of those people who helped and collaborated with me during last years.

First of all, I would like to express my sincere gratitude to my two supervisors and as well as mentors: Prof. Manuel Martínez and Prof. Genaro Saavedra. I would not have been able to reach as far as here and accomplish my doctoral program, without their unstinted support and great teachings. They invested in my potentials and backed me up all along the way without any doubt, and sometimes gave me their hearty and sincere advice to resolve the critical problems what I faced. Their wisdom and enthusiasm for knowledge have had a great impact on me, and I will never forget it. In addition, I will remember their teachings throughout my life and engrave them in my heart, and use them as milestones and indicators for the future.

Meanwhile, I would not have been able to participate and start the doctorate program without any helps and supports from Prof. Byung-Gook Lee, Dr. Donghak Shin, and Prof. Myungjin Cho. They have opened and guided me the way forward, and have given me this tremendous opportunity with their full material and emotional support. Moreover, it is no exaggeration to say that their academic achievements and subject of research not only have had a great influence in determining the direction of my research, but also inspired me what to do afterwards. I would like to express my cordial appreciation and deep gratitude for their support and help.

Also, I would like to express my sincere boundless gratitude for other professors from Department of Optics in the University of Valencia (Spain), especially, thank to Prof. Juan Carlos Barreiro, Prof. Pedro Andrés, and Prof. Amparo Pons. They gave me realistic advices and as well as shared their various experiences as a scholar who has been researching for a long time.

Now, I would like to express many thanks to the members of my group *3D Imaging & Display*, those who I worked and collaborated with, and helped me anytimes, in any way possible, by the best of their abilities.

Gracias a Adrián por muchas cosas. Hemos colaborado durante los últimos 4 años, y también me serviste de ayuda para comprender el concepto de la Óptica y cómo pudimos progresar en las investigaciones. Realmente me gustó y me entusiasmó colaborar en nuestras investigaciones comunes durante esos períodos contigo.

Gracias a Jorge que me ayudaste en cualquier momento y muchas veces en el ámbito personal. También, me enseñaste sobre la teoría básica de la Óptica y aún más allá. Si no hubiera recibido tu ayuda, tendría una situación realmente difícil e incluso podría haberme enfrentado a muchos problemas durante los últimos años.

Mucho ánimo a mis compañeros Emilio, Anabel, Ángel, Gabriele, Amir, Nico, Chemek, y Alejandro. Diría que los últimos años fueron muy feliz para estudiar y estar con vosotros por todos juntos. Ahora os toca tomar el relevo y ser los responsables de sacar adelante de investigación de nuestro grupo 3DDI. Estoy seguro de que vosotros lo haréis muy bien y terminaréis vuestro estudio sin problema y genial.

Gracias a Ana que me ayudaste mucho cuando llegué a España y comencé el doctorado. Si no hubiera recibido tu ayuda en ese tiempo, sería muy difícil y no podría establecerme adecuadamente. Espero que bendiga tu futura vida en Estados Unidos con Benja y tu familia, también tus perritos.

No puedo faltar aquí un párrafo para mi compañero de trabajo, y también, un compañero de gimnasio, Héctor. Gracias a Héctor por tu sincera amistad e incluso investigaciones comunes en los últimos años. Espero que podamos ser amigos cercanos durante muchos años, y también deseo que hagamos las colaboraciones en nuestras investigaciones comunes en el futuro.

Thanks to Kotaro, 병우, 윤희, 영준, and 기옥 during my short stay in Hankyong National University in South Korea. Especially, it was great moment for me to collaborate and study with Kotaro, who is also taking part in Ph.D program. I look forward to waiting for our further collaboration in the near future. I hope that all of you could finish your study and researches successfully, and also I wish you all the best in future.

Sin duda, durante la realización de este doctorado, he podido conocer a mucha gente maravillosa. No puedo olvidarme a todos los compañeros del departamento que han hecho que la hora del café o la hora de la comida fueron los momentos agradables. Entre ellos están, Braulio, Diego, Vicent, Lluís, David de Tenerife, Sento, Ade, Esther, Oleksii de Ukraine, Luis de Costa Rica, Aktham de Jordania, José Ángel, Alejandro, José María, Abril, Ivan & Isabel, y Héctor Ramírez de México. Especialmente, me gustaría expresar mi sincero agradecimiento a Braulio, quien ha sido muchos los momentos que hemos compartido juntos, y también con su sincera amistad. Mientras tanto, gracias por el grupo deportivo de equipo UV del Fútbol 7 y otro equipo del fútbol sala de Burjassot, para enseñar y aconsejar a mí durante de los últimos años.

Pero entre todos, gracias a la Universitat de València, por la financiación otorgada durante de estos años en el marco del programa “Atracció de Talent”. También, gracias a l@s profesores de la escuela oficial de idiomas (EOI Valencia) por su curso de español con pasión.

Ahora, me gustaría agradecer a numerosas personas con las que he compartido gratos momentos durante de los últimos años en Valencia. Entre ellos están, Maite, Ángel & Carina, David, Ana, Lorena, Carlos, Sonia, Laura, Ani, Yao, Rennei, Presley, Julio de la peluquería, Meddy de Marruecos, Lluís de Barcelona, Pedro, y Manuel. Con especial mención a Maite por su ayuda cualquier momento y muchas veces en el ámbito personal, y también por su sincera amistad. Perdonad si me dejo a alguien, pero es que sois muchísimos. En todo caso, gracias por todos los momentos que hemos compartido y por ayudarme en los momentos durante de los últimos años. Os echaré de menos, chic@s. Nos vemos muy prontito.

또한, 스페인이라는 타지에서 만나, 힘든 시기에 여러모로 물심양면 도움주신 많은 한국인분들, 특히 수경 누나 & 태근 큰형님, 현진 선생님 & 호재 형님, 미나 선생님 & Pau 형님, 김진학 한인회 회장님, 순천 & 재희, 선미 & Jorge, 이석 형님 & 나리, 준석, 동균, 수경, 필제 형님, 인기, 지은, 한결, 중지, 정원 누나, 경진, 인수 & Julián, 지수, 용진, 그리고 또한 혹여나 언급이 되지 않은 수많은 분들께도 진심으로 감사하다는 말씀을 전해드리고 싶습니다.

한편, 스페인으로 유학 오는데 정말 큰 도움주신 동록 외삼촌께 진심 어린 감사의 뜻을 전하고 싶습니다. 스페인으로의 유학을 준비하는 과정에서 성심성의껏 도와주셔서, 그때 직면했던 크나 큰 역경을 문제없이 잘 극복하고 해결하여, 마침내 위기를 무사히 넘길 수 있었던 것 같습니다. 제 꿈을 펼칠 수 있도록 도움을 주셔서 정말 감사합니다, 삼촌.

마지막으로, 저를 언제나 믿고 지지해주시며, 유학이라는 크나큰 결정을 한 점 망설임 없이 허락해주시고, 존중해 주시며, 평생토록 든든한 저의 버팀목이 되어준 우리 가족, 아버지, 어머니, 동생, 그리고 홀로 기나 긴 인고의 시간을 멀리서 묵묵히 견디고 기다려주며, 힘과 용기를 북돋아주고, 언제나 늘 포근한 안식처가 되어준 나의 아내, 혜현이에게 진심어린 사랑과 깊은 감사의 마음을 전하며, 이 논문을 바칩니다.

Abstract

During the last century, various technologies of 3D image capturing and visualization have spotlighted, due to both their pioneering nature and the aspiration to extend the applications of conventional 2D imaging technology to 3D scenes. Besides, thanks to advances in opto-electronic imaging technologies, the possibilities of capturing and transmitting 2D images in real-time have progressed significantly, and boosted the growth of 3D image capturing, processing, transmission and as well as display techniques. Among the latter, integral-imaging technology has been considered as one of the promising ones to restore real 3D scenes through the use of a multi-view visualization system that provides to observers with a sense of immersive depth. Many research groups and companies have researched this novel technique with different approaches, and occasions for various complements.

In this work, we followed this trend, but processed through our novel strategies and algorithms. Thus, we may say that our approach is innovative, when compared to conventional proposals. The main objective of our research is to develop a technique that allows recording and simulating the natural scene in 3D by using several cameras which have different types and characteristics. Moreover, we provide a volumetric scene which is restored with great similarity to the original shape, through a comprehensive 3D monitor. Our Proposed integral-imaging monitor shows an immersive experience to multiple observers.

In this thesis we address the challenges of integral image production techniques based on the computerized 3D information, and we focus in particular on the implementation of full-parallax 3D display system. We have also made progress in overcoming the limitations of the conventional integral-imaging technique. In addition, we have developed different refinement methodologies and restoration strategies for the composed depth information. Finally, we have applied an adequate solution that reduces the computation times significantly, associated with the repetitive calculation phase in the generation of an integral image. All these results are presented by the corresponding images and proposed display experiments.

Resumen

Durante el siglo pasado, las tecnologías de captura y visualización de imágenes en 3D han destacado con fuerza, debido tanto a su carácter pionero como por el atractivo de la aspiración de extender a escenas 3D las aplicaciones de la tecnología de imagen convencional. Gracias a los avances en las tecnologías opto-electrónicas de imagen, las posibilidades de capturar y transmitir imágenes 2D en tiempo real han progresado significativamente e impulsado el crecimiento de las técnicas de captura, procesado, transmisión y display de imágenes 3D. Entre estas últimas, la tecnología de imagen integral ha sido una de las más prometedoras para reproducir las escenas reales 3D mediante el uso de un sistema de visualización multi-perspectiva que proporciona a los observadores una sensación de profundidad inmersiva. Muchos grupos de investigación y compañías han investigado esta técnica novedosa desde enfoques diferentes, y en muchas ocasiones complementarios.

En esta Tesis, nos adherimos a esta tendencia, pero hemos abordado nuestro objetivo de una manera distinta, basada fundamentalmente en el desarrollo y optimización de nuevos algoritmos. Nuestro enfoque es innovador, al compararlo con las propuestas convencionales. El objetivo principal de nuestra investigación es desarrollar una técnica que permita registrar la escena natural en 3D mediante el uso de varias de cámaras de diferente tipo, y mostrar una escena volumétrica restaurada con gran semejanza con la original a través de un monitor integral 3D adaptado para presentar una experiencia inmersiva a múltiples observadores en conjunto.

En esta Tesis abordamos los desafíos de las técnicas de producción de imagen integral basadas en información computarizada en 3D, y nos concentramos en particular en la implementación del sistema de display 3D de paralaje completo. También hemos avanzado en la superación de las limitaciones de la técnica convencional de imagen integral. Además, hemos desarrollado diferentes metodologías de refinamiento y estrategias de restauración para la información de profundidad. Por último, hemos obtenido una solución adecuada que permite aminorar significati-

vamente los tiempos de computación asociados al cálculo repetitivo requerido en fase de generación de una imagen integral. Naturalmente, todos estos resultados están respaldados por los correspondientes experimentos de imagen y display 3D.

Resumen extendido

Durante el presente siglo, las técnicas de registro y reproducción de imágenes tridimensionales han destacado por su capacidad de capturar, procesar y mostrar la información especial completa de escenas reales o sintéticas. Como prueba de ello, en la actualidad se desarrollan, a nivel mundial, investigaciones y actividades sobre tecnologías 3D en muchos campos diferentes. De hecho, se ha presentado una inmensa cantidad de artículos en revistas y conferencias sobre imágenes y visualización en 3D, y se han realizado también grandes esfuerzos en investigación y desarrollo tanto a nivel gubernamental como industrial. Las aplicaciones en esta área incluyen la fabricación de dispositivos de consumo masivo, la seguridad y defensa de territorios y/o personas individuales, la automatización de máquinas, las aplicaciones biomédicas e, incluso, el entretenimiento. No nos equivocamos si anunciamos que la década actual será la década de las imágenes en 3D.

Esta tesis está organizada en seis capítulos. El Capítulo 1 comienza con una breve crónica histórica de la obtención y reproducción de imágenes 3D y luego se presentan motivadamente los objetivos del trabajo. En ese contexto, se señala que, en la actualidad, la mayoría de las técnicas de captura y generación de imágenes tridimensionales se basan en la estereoscopía. La estereoscopía se define como cualquier tecnología que permita una percepción de profundidad para observadores binoculares. La estereoscopía convencional opera con dos imágenes que se presentan en canales separados al ojo izquierdo y derecho del observador. Esta pareja de imágenes tiene puntos de vista (o perspectivas) ligeramente diferentes. Así, el observador capta esta paralaje binocular a través de ambos ojos y deja que el cerebro obtenga la percepción de profundidad debido a su disparidad visual. Es el cerebro el que también determina la distancia y lo lejanos que están los objetos entre sí a través de la magnitud de la disparidad entre las dos imágenes. Desde un punto de vista histórico, se indican las fuentes que citan ya a Euclides como conocedor de la percepción de profundidad binocular, siendo de los primeros en reconocer en el siglo III que la percepción de profundidad se obtiene cuando cada ojo recibe una de dos imágenes diferentes del mismo objeto al mismo tiempo. En el siglo XVI, Leonardo da Vinci estudió aspectos de la óptica, incluida la anatomía del ojo y el reflejo de la luz, e incluso intentó explicar la percepción estereoscópica de profundidad, mencionando que un objeto dado ocluye diferentes partes del fondo cuando se observan con el ojo izquierdo en comparación con el ojo derecho. Ya en el siglo XIX, el científico británico Charles Wheatstone inventó el primer instrumento diseñado para observar esas imágenes y producir una sensación 3D, a partir de su ”Estereoscopio de espejo reflectante” formado por dos espejos centrados a 45º de cada ojo del observador. A pesar de funcionar con láminas dibujadas, este dispositivo inspiró a muchas compañías fotográficas posteriores a abrir un nuevo mercado comercial basado en la fotografía estereoscópica y el estereoscopio, alcanzando su auge en el siglo XX. En cualquier caso, la tecnología de entonces solo podía proporcionar una buena experiencia visual 3D mediante el uso de anteojos especiales o la utilización de dispositivos adicionales para el observador binocular. De entre los que tuvieron más difusión, los anaglifos fueron originalmente los más usados para la codificación/descodificación de estereogramas. Los anaglifos contienen dos imágenes de color diferente, en las que dichas imágenes tienen un contraste complementario (como los colores rojo-cian, verde-magenta o azul-amarillo). Cuando se observa a través de las gafas de anaglifo codificadas por colores, cada una de las dos imágenes es percibida por cada ojo por separado y el cerebro combina ambas imágenes dispares integrándolas en una imagen estereoscópica. Esta sencilla técnica presentaba, sin embargo, inconvenientes graves a la hora de reproducir objetos 3D en color, causando fuertes distorsiones cromáticas en la percepción de los objetos. Por ello, fue siendo sustituida progresivamente por otros modos de codificación/descodificación basados en propiedades de la luz (gafas pasivas) o del sistema visual humano (gafas activas). Las gafas pasivas usan la polarización de la luz (a la que el sistema visual humano es básicamente insensible) para, mediante el uso de filtros polarizadores, crear una ilusión de imágenes en 3D al restringir la escena que se ajusta con su respectiva información a cada ojo de manera distinta, de modo que cada ojo individual vea solo la imagen hecha para esa perspectiva, y viceversa. Por otro lado, las gafas activas presentan la imagen destinada al ojo izquierdo mientras bloquean la vista del ojo derecho, y luego presentan la imagen del ojo derecho mientras bloquean la vista del ojo izquierdo. Este proceso se repite secuencialmente de modo muy rápido (por encima de la frecuencia crítica de fusión del sistema visual humano) para que las interrupciones no afecten a la fusión percibida de las dos imágenes en una sola imagen 3D. Ambos sistemas presentan ventajas e inconvenientes. Por ejemplo, el filtro polarizador de gafas pasivas no puede bloquear cada imagen clasificada correctamente en algunas situaciones específicas. Así, si hay algún objeto oscuro al lado de algo brillante en la escena mostrada, un ojo puede notar la luz que está destinada al otro ojo y, por lo tanto, aparecen algunas áreas interferidas. Por otro lado, las gafas activas resuelven esta diafonía 3D o efecto fantasma correctamente. Sin embargo, las gafas pasivas son económicas y no se requieren baterías, por lo que presentan un peso mucho más reducido. Además, las gafas pasivas carecen de efecto de parpadeo residual, por lo que este tipo de gafas proporcionan menos molestias en su uso prolongado. Ambas técnicas presentan una deficiencia fundamental en su propósito de proporcionar una verdadera inmersión 3D al observador: sólo proporcionan una única perspectiva de la escena, uniforme para todos los observadores incluso cuando se colocan en diferentes posiciones de la escena visualizada. Por lo tanto, la escena mostrada parece ser demasiado artificial y estar muy alejada de la experiencia real. En su versión más sofisticada, ambas tecnologías se pueden integrar en sistemas de reproducción individualizada a través del uso de displays de soporte frontal (HMD), montado en la cabeza o como parte de un casco, que disponen de una pequeña pantalla para proporcionar la escena para uno o ambos ojos directamente. Los HMD tienen una gran demanda en la actualidad por sus aplicaciones en la seguridad y la defensa, el entrenamiento militar o deportivo, la terapia perceptual e incluso el entretenimiento por la experiencia 3D del consumidor general. Los dispositivos HMD recientes disponen de un sensor giroscópico y reaccionan al movimiento de la cabeza del observador, pudiendo modificar en tiempo real la perspectiva presentada al observador, pero aún son lentos para reproducir la verdadera sensación 3D de la vida real. Además, la ergonomía de los HMD, que los observadores se ponen obligatoriamente en sus cabezas para ver la escena 3D, es muy reducida debido a su tamaño voluminoso y su peso considerable. Además, todos estos sistemas, , producen un efecto colateral que impide su uso prolongado en la mayor parte de los observadores. Tanto las gafas de anaglifo como las gafas pasivas, las activas y los HMD se basan en la generación artificial de la sensación de profundidad a partir de la disparidad de las imágenes proporcionadas a cada ojo del observador mientras que se ha de mantener constantemente el enfoque del sistema visual sobre el display (ilusión estereoscópica). Este modo de funcionamiento proporciona, en el mejor de los casos, fatiga visual asociada al conflicto convergencia/acomodación inherente a la observación a través de estos dispositivos. En la mayor parte de los casos, tras un uso prolongado de estos sistemas, el observador deja de percibir la sensación tridimensional de la escena. Un enfoque alternativo para generar las escenas 3D sin necesidad de ninguna herramienta adicional o equipo portátil es la autoestereoscopía, en la que se proporciona al observador la reproducción 3D a simple vista directamente. La idea principal de esta técnica es proporcionar numerosas vistas en perspectiva desde una única pantalla, dentro de un rango determinado, permitiendo, además, que varios observadores pueden ver la escena 3D estereoscópica con sus propios ojos, manteniendo cada uno perspectivas distintas de la misma. Una de los métodos utilizados para generar estos displays autoestereoscópicos se basa en la denominada fotografía integral, propuesta por el Premio Nobel de Física G. Lippmann en 1908. La idea planteada consistió en capturar la escena 3D a través de una matriz de microlentes (o estenopes, en propuestas anteriores) y reconstruirla con el mismo dispositivo, sustituyendo el sensor (película fotográfica, en aquel tiempo) por la imagen registrada (entonces película revelada) e iluminando en sentido contrario el conjunto. Sobre el medio de registro se obtiene un conjunto (imagen integral) de imágenes pequeñas diferentes (imágenes elementales), cada una con información de perspectiva diferente, y se restablece la escena registrada en orden inverso al de la etapa de captura. Tras la publicación de los resultados de Lippmann, muchos equipos de investigación se involucraron en su desarrollo posterior y se convirtieron en un catalizador para la investigación sobre la captura y regeneración de imágenes autoestereoscópicas. Sin embargo, su impacto tecnológico fue reducido por la inmadurez de las técnicas de registro disponibles, que no se adaptaban fácilmente al doble proceso de captura/reproducción de la fotografía integral. En las últimas décadas, gracias al avance en los sensores electrónicos de imagen, se han implementado varias de estas propuestas para capturar y transmitir imágenes en tiempo real y varias empresas, finalmente, han comercializado productos tanto de aplicación técnica como de consumo masivo basados en la técnica de Lippmann. Estas cámaras integrales capturan la información espacio-angular de una escena, a partir de la cual se pueden realizar diversas operaciones de procesado. En particular, estas cámaras pueden componer un mapa de profundidad de los objetos registrados a partir de la imagen capturada en un solo disparo. Esto se consigue de modo pasivo empleando iluminación convencional, a diferencia de otros dispositivos de captura de mapas de profundidad en los que se requiere generar una iluminación codificada espacial y/o temporalmente de la escena (como en las cámaras con iluminación estructurada o los sensores de tiempo de vuelo). Otra dificultad técnica en la que la fotografía integral se ha mostrado especialmente útil se refiere a la representación de la información 3D obtenida a partir de los mapas de profundidad. A pesar de que existen muchas técnicas excelentes de detección de profundidad y un gran avance de las tecnologías digitales, los sistemas de visualización convencionales (como televisión, monitor, teléfono móvil o incluso tabletas digitales) no pueden proporcionar al observador la escena 3D real como la información volumétrica original. Esto se debe a que estos sistemas de visualización solo pueden mostrar simultáneamente proyecciones 2D (vistas o perspectivas) de la escena registrada. Sin embargo, el objetivo último de los sistemas que estamos estudiando es reproducir y mostrar la información 3D tal como es realmente y proporcionar una réplica de la escena 3D original a los observadores utilizando la información 3D digitalizada y procesada en la etapa de captura. La segunda etapa del proceso descrito por Lippmann ha proporcionado la clave para poder desarrollar sistemas de display 3D en los que la información volumétrica capturada (por cualquier dispositivo que pueda generar un mapa de profundidad, estén o no basados en la fotografía integral) se presenta al observador tal y como se distribuía en la escena real original. Éste es el tipo de display al que se han adaptado todos los resultados de esta tesis para su proyección realista, accesible a múltiples observadores simultáneos, con paralaje dinámica, y sin requerir ningún equipo portátil adicional o anteojos para observar la escena 3D real restaurada. Todo esto lleva naturalmente a la definición de los objetivos de esta tesis. En este trabajo se abordan varios de los desafíos de las técnicas y los algoritmos de producción de imágenes integrales y se centra, en particular, en la composición de diferentes sistemas de captura y visualización 3D de paralaje completo. Así, se realizan diferentes propuestas para superar algunas limitaciones de la técnica de imagen integral convencional y se discuten varias metodologías de refinamiento en la extracción de la información 3D a partir de la información registrada por diferentes cámaras/sensores de profundidad. Además, se proponen diferentes estrategias de generación de imágenes integrales, a partir de registros de diferentes tipos de sensores de profundidad, para la proyección 3D realista en un monitor basado en imagen integral. Todas las propuestas se han validado tanto desde el punto de vista de la implementación computacional como su verificación experimental sobre los monitores integrales disponibles en el Laboratorio de Imagen y Display 3D de la Universitat de València.

El capítulo 2 ofrece una base teórica de las aportaciones principales de esta tesis. En la primera sección, se describe detalladamente la técnica de imagen integral (InI) y el problema pseudoscópico asociado. InI es una muy prometedora tecnología de visualización y captura en 3D que ofrece un variación continua del punto de vista, paralaje completa y vistas a todo color para múltiples observadores simultáneos. Pero, entre todos, el mérito principal de InI es transcribir la información espacial y angular de los rayos que proceden de la escena 3D al mismo tiempo. Basado en el enfoque ya citado de la fotografía integral propuesta por G. Lippmann, InI es capaz de grabar la escena natural en 3D y mostrar la escena restaurada usando una matriz de elementos dióptricos (p.e., microlentes). La presentación de este mismo registro (originalmente en forma de película fotográfica revelada y sobre un display digital en la actualidad) frente a la misma matriz de captura regenera una imagen flotante y produce una reconstrucción 3D de la escena original capturada. Sin embargo, a pesar de la eficacia de esta técnica de reproducción de escenas 3D, es importante señalar que la imagen capturada mediante la técnica InI no es directamente proyectable si se busca una percepción correcta de la profundidad relativa de los objetos en la escena. El problema para dicha reconstrucción fiel es que la reproducción directa genera una imagen 3D pseudoscópica, en la que los objetos más cercanos al observador son los que aparecía más lejanos al sistema de captura, y viceversa. Este efecto se produce por el cambio en el sentido de propagación de la luz en la secuencia del proceso de registro y de visualización. Nótese que en la fase de captura, los rayos de luz dispersados del objeto 3D pasan a través de cada elemento de la matriz y cada uno compone una llamada imagen elemental (EI). En cambio, en la fase de visualización, el objeto 3D es reconstruido por las EI con el mismo conjunto de elementos pero iluminados en sentido contrario. Así, desde la escena visualizada, los observadores visualizarán la parte posterior de la imagen 3D reconstruida en la dirección del objeto 3D a la matriz de captura/reproducción. Este efecto implica que la escena 3D visualizada final tiene una profundidad invertida en 180°. Dicho de otro modo, un objeto 3D capturado cerca de la matriz reconstruye la escena 3D más cerca de ésta, y un objeto más alejado reconstruye la escena más lejos de la matriz, respectivamente. Por lo tanto, las EI sin ningún proceso adicional o método de transformación no pueden evitar el problema de la pseudoscopia de la escena reconstruida. En los capítulos siguientes de la tesis se presentan con detalle algunos métodos para resolver estos problemas. Todavía en el Capítulo 2, se pasa a presentar con detalle los diferentes tipos de cámaras que se usan en los experimentos desarrollados. La metodología sobre cómo grabar y reproducir fielmente una escena natural en 3D es una tarea de investigación de las más destacadas durante las últimas décadas, como parte natural del proceso en el desarrollo de la era digital. De hecho, la captura de la información 3D no puede basarse en el uso de una cámara fotográfica convencional, ya que ésta transcribe y guarda la escena natural capturada en la información 2D del plano de registro, perdiendo la información volumétrica 3D original. Además, la información 2D registrada no puede presentar e interpretar adecuadamente las áreas ocluidas u ocultas e, incluso, tampoco puede observar las partes veladas en diferentes vistas en perspectiva. Por el contrario, sólo cuando se registra la información 3D completa se superan estas limitaciones. La última sección del capítulo se dedica a la descripción de las técnicas empleadas en la tesis para la generación, manipulación y reproducción de esta información 3D a partir de la generación digital de nubes volumétricas de puntos.

En el Capítulo 3, se describen varias metodologías para componer una imagen integral a partir de los datos 3D computarizados anteriores (nube de puntos), de una manera más eficiente y precisa que en los métodos estándar que se utilizan actualmente en InI. En la primera sección, se presentan nuevos métodos para componer la imagen integral utilizando una nube de puntos con diferentes enfoques.

En este sentido, se proponen dos metodologías sobre cómo construir la imagen integral utilizando una nube de puntos y una matriz virtual de estenopes (VPA). El primer método consiste en situar el VPA dentro (o cerca) de la nube de puntos y realizar el esquema de proyección desde cada punto a todos los estenopes uno tras otro. La segunda estrategia propuesta es disponer el VPA lejos de la nube de puntos y recoger las perspectivas y, posteriormente, calcular la imagen integral. A continuación, se presenta un estudio comparativo de ambos algoritmos, mostrando sus ventajas y desventajas relativas e identificando el tipo de escenarios en los que el uso de una u otra de las técnicas es más adecuado y eficiente. En la siguiente sección del Capítulo 3, se explica con detalle la técnica de aceleración de algoritmos a través de paralelización por hardware y se presenta cómo se utiliza en la tesis. De hecho, nos enfrentamos al defecto crítico conocido del esquema de cálculo repetitivo pesado en el procedimiento de generación de imágenes integrales, que lo hace inviable con el procesado en tiempo real a frecuencia estándar de video. Una de las técnicas de aceleración de hardware más habituales es la conocida informática acelerada por unidades específicas de procesado gráfico (GPU). La técnica de aceleración por GPU se basa en el uso de hardware especialmente diseñado para procesar rápidamente grandes cantidades de datos, que se utilizan para realizar cálculos pesados de manera más eficiente de lo que es posible en las unidades de procesado central (CPU) de uso general. Así, la GPU se usa principalmente en las partes de los códigos que requieren mucho tiempo de cómputo al iterarse de modo intensivo operaciones elementales sobre datos independientes, y el resto de las aplicaciones se ejecutan simultáneamente en la CPU. La adaptación de esta técnica al procesado en InI mejora el rendimiento drásticamente. En la tesis se explica detalladamente cómo se aplica en nuestro caso y se presenta el resultado de la comparación entre el rendimiento usando sólo CPU y con la aceleración a través de GPU. Finalmente, en la última sección, se describe detalladamente el procedimiento seguido para preparar la información procesada para su uso en un sistema de visualización InI.

En el Capítulo 4 se presentan los métodos propuestos sobre cómo mejorar la calidad de los datos 3D, cómo componer un mapa de profundidad sin distorsiones a partir de una imagen integral capturada, y cómo recuperar las áreas perdidas de las nubes de puntos. Como se mencionó anteriormente, en esta tesis se explota la información 3D generada digitalmente en forma de nube de puntos para componer una imagen integral que pueda reproducir la escena capturada con información espacial completa. En consecuencia, una nube de puntos especialmente densa no solo ayuda a crear una buena calidad de la imagen integral, sino que también ayuda a proporcionar una escena 3D más inversiva a los observadores. Sin embargo, los mapas de profundidad generados presentan ciertos defectos debidos a diferentes causas. Algunos de ellos provienen de la propia limitación de la cámara 3D utilizada en la captura de la escena original o de la pérdida de información de las áreas ocluidas y/o ocultas en función de la posición de la cámara. Así, en la primera sección de este capítulo, se introduce una nueva técnica para restaurar las regiones de profundidad ambiguas o perdidas e, incluso, las áreas ruidosas del mapa de profundidad capturado. Nos centramos aquí en sensores de profundidad por infrarrojos (IR). De hecho, la técnica de detección de profundidad IR es una de las más ampliamente utilizadas, a pesar de que dichas cámaras tienen varios inconvenientes. Uno de los más destacados es que adquieren las imágenes de profundidad con cierto ruidos y/o "agujeros" debido a sus propias limitaciones y/o a factores externos (como la iluminación de las muestras que puede interferir en los haces IR utilizados en la determinación del mapa de profundidad). El objetivo principal en esta parte de la tesis es recuperar y mejorar la calidad de la imagen de estos mapas de profundidad, adoptando un método adaptado de filtrado de áreas vacías y/o ruidosas con muy buen rendimiento de restauración y robustez. Esta implementación se valida también con resultados experimentales que permiten establecer la comparación entre los métodos convencionales y la propuesta original de este trabajo. En la siguiente sección, se introduce una nueva metodología sobre cómo componer un mapa de profundidad ultradenso a partir de una imagen integral capturada en una sola toma. De hecho, como mencionamos anteriormente, las cámaras de InI tienen capacidades únicas de recolección de la información espacio-angular 3D de las escenas. En la tesis modificamos y mejoramos una estrategia de estimación de profundidad ya publicada para poder compensar ciertas distorsiones de imagen que aparecen en la imagen integral, combinando el método original con una técnica de calibración específica para este tipo de cámaras y validándolo experimentalmente. Con ello, se pudo generar un mapa de profundidad sin distorsiones de una escena real implementada en el laboratorio. Finalmente, se concluye este capítulo presentando un método de registro de datos 3D para componer nubes de puntos ultradensas como combinación de dos nubes dispersas. Las técnicas de registro de imágenes es una tarea rutinaria que superpone dos o incluso más imágenes de la misma escena, que fueron capturadas desde diferentes perspectivas por varios sensores y/o cámaras. Entre otros, el algoritmo iterativo de búsqueda del punto más cercano (ICP) es una de las técnicas más utilizadas para fusionar los pares de datos 2D y/o 3D. El algoritmo ICP tiene como objetivo encontrar el punto más cercano a un punto dado en una entidad geométrica, y calcula el desplazamiento entre los conjuntos de datos utilizando un procedimiento de refinamiento iterativo. En este trabajo se aplica el algoritmo ICP básico para superar las limitaciones de los sistemas de visión monocular y, en particular, completar las áreas ocluidas y/u ocultas sobre la nube de puntos compuesta. Esto ocurre porque la inherente pérdida de áreas superpuestas sobre la línea de visión de una cámara monocular puede superarse por el empleo de vistas múltiples que amplían el FOV y recuperan la información ocluida al complementarse entre sí. En la memoria se establecen dos composiciones de cámara estéreo diferentes para la generación de nubes de puntos compuestas: una es un sistema híbrido de dos cámaras de profundidad y la otra es un sistema estéreo de cámaras InI. En el sistema de cámaras 3D estéreo híbrido se utilizaron dos cámaras de detección de profundidad basadas en tecnologías diferentes, a saber, una cámara de profundidad de proyección de patrón IR estructurado y otra es una cámara de medida de tiempo de vuelo (ToF). Ambas cámaras de detección de profundidad tienen características totalmente diferentes y, por tanto, estos factores diferentes deben homogeneizarse para crear un conjunto de nubes de puntos uniforme y evitar irregularidades visuales en la escena 3D reconstruida. Por otro lado, en el sistema de cámara estéreo-InI, se explota una cámara de imagen integral comercial con control deslizante para capturar una misma escena desde dos posiciones diferentes. En ambos casos, la configuración estéreo permitieron recuperar algunas áreas ocluidas y perdidas de las nubes de puntos individuales, obteniéndose en el laboratorio reconstrucciones mejoradas de escenas 3D reales.

El Capítulo 5 consiste en la recopilación de las principales publicaciones académicas que constituyen la base para el desarrollo de esta tesis. Cada artículo proporciona enfoques diferentes y distintos sobre cómo componer y refinar los datos 3D densos adquiridos por los diferentes tipos de cámaras y exponen también las técnicas y soluciones originales sobre cómo proporcionar una experiencia 3D más inmersiva a los observadores a través del use del monitor de imagen integral propuesto. Finalmente, el Capítulo 6 concluye esta tesis con una exposición de los logros alcanzados y con algunos comentarios sobre nuestras posibilidades futuras de investigación.

List of Publications

This dissertation is based on the following 6 main articles (Papers I-VI). Other research results from various conference proceedings and patents are also enumerated.

Main articles

[I] **S. Hong**, D. Shin, B-G. Lee, A. Dorado, G. Saavedra, and M. Martínez-Corral, "Towards 3D Television Through Fusion of Kinect and Integral-Imaging Concepts", J. Disp. Technol. 11, 894-9 (2015).

[II] **S. Hong**, A. Dorado, G. Saavedra, J. C. Barreiro, and M. Martínez-Corral, "Three-Dimensional Integral-Imaging Display From Calibrated and Depth-Hole Filtered Kinect Information", J. Disp. Technol. 12, 1301-8 (2016).

[III] **S. Hong**, G. Saavedra, and M. Martínez-Corral, "Full parallax three-dimensional display from Kinect v1 and v2", Opt. Eng. 56, 041305 (2017).

[IV] **S. Hong**, A. Ansari, G. Saavedra, and M. Martínez-Corral, "Full-parallax 3D display from stereo-hybrid 3D camera system", Opt. Laser. Eng. 103, 46–54 (2018).

[V] N. Incardona, **S. Hong**, M. Martínez-Corral, and G. Saavedra, "New Method of Microimages Generation for 3D Display", Sensors 18, 2805 (2018).

[VI] **S. Hong**, N. Incardona, K. Inoue, M. Cho, G. Saavedra, and M. Martínez-Corral, "GPU-accelerated integral imaging and full-parallax 3D display using stereo–plenoptic camera system", Opt. Laser. Eng. 115, 172–8 (2019).

Patents

[VII] H. Navarro, J. Sola-Pikabea, **S. Hong**, J. C. Barreiro, G. Saavedra, and M. Martínez-Corral, "Method and device for depth detection using stereo images", PCT/EP2016/075714 (2016).

Conference proceedings

[VIII] **S. Hong**, A. Dorado, G. Saavedra, M. Martínez-Corral, D. Shin, and B-G. Lee, "Full-parallax 3D display from single-shot Kinect capture", Proc. SPIE 9495, 94950E (2015).

[IX] **S. Hong**, A. Dorado, G. Saavedra, J. Sola-Pikabea and M. Martínez-Corral, "Full-parallax 3D display from the hole-filtered depth information", 3DTV-CON art. no. 7169355, 1-4 (2015).

[X] **S. Hong**, G. Saavedra, and M. Martínez-Corral, "Full-Parallax immersive 3D display from Depth-Map cameras", Workshop on Information Optics art. no. 7745584, 1-3 (2016).

[XI] **S. Hong**, A. Ansari, G. Saavedra, and M. Martínez-Corral, "Integral-Imaging display from stereo-Kinect capture", Proc. SPIE 10219, 102190K (2017).

Acronyms

3D	Three-dimensional
HMD	Head-mounted display
IP	Integral photography
InI	Integral imaging
2D	Two-dimensional
PS	Pseudoscopic
EIs	Elemental images
OS	Orthoscopic
4D	Four-dimensional
MLA	Microlens array
FOV	Field of view
IR	Infrared
Kv1	Kinect v1
Kv2	Kinect v2
ToF	Time-of-Flight
VPA	Virtual pinhole array
DOF	Depth of field

GPUs	Graphics processing units
CPUs	Central processing units
GPGPU	General-purpose computing on GPU
LCD	Liquid crystal display
ppm	Pixels per millimeter
JBF	Joint-bilateral filter
BF	Bilateral filter
ICP	Iterative closest point

Contents

Chapter 1

Introduction

During the present century, three-dimensional (3D) imaging techniques have been spotlighted due to their merits of capturing, processing, and displaying 3D scenes. As proof of this, widespread international researches and activities on 3D technologies are performed and applied in many different research fields. In fact, there are a large number of journal and conference papers on 3D imaging and display, as well as big efforts in research and development on government, industry, military, and various different laboratories. This topic is aimed to be exploited for broad applications including manufacturing, security and defense, surveillance, machine automation, biomedical applications, and even entertainment. It will not be wrong to say that current decade will be the decade of 3D imaging.

Most of the 3D imaging techniques is based upon stereoscopy. Stereoscopy is defined as any technology that enables an illusion of the depth in a display to binocular observers. Conventional stereoscopy manipulates with two offset images to the left and right eye of the observer. A noteworthy feature is that the pair of photo has slightly different viewpoints (or perspectives). Thus, the observer accommodates this binocular parallax through the both eyes and let the brain get the depth perception because of their visual disparity. The brain also determines the depth distance and how far objects are away from each other by the amount of disparity between two images.

In [1], Kheirandish mentioned that stereoscopic depth perception was early assumed by the ancient Greeks. Sir David Brewster mentioned that Euclid knew of the binocular depth perception. As stated in his book [2], in 280 A.D., Euclid was the first to recognize that the depth perception is obtained when each eye receives one of two dissimilar images of the identical object at the same time. However,

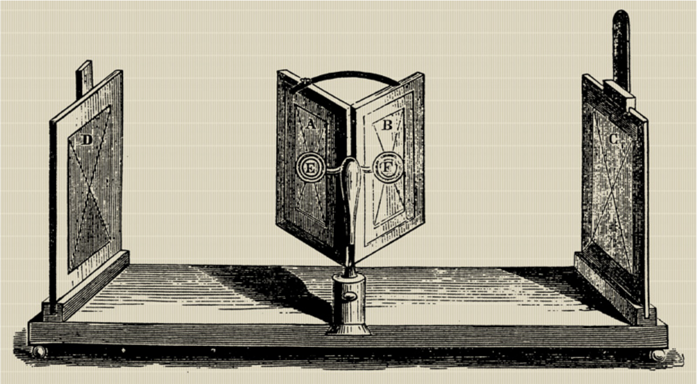


Figure 1.1: The diagram of stereoscope from Wheatstone's paper published in the Transactions of the Royal Society of London, 1838.
source by: `https://commons.wikimedia.org/wiki/Category:Wheatstone_stereoscopes#/media/File:Charles_Wheatstone-mirror_stereoscope_XIXc.jpg`, ©Wikimedia Commons

many people tackled Brewster's statement about the achievement of Euclid [3], but, it is an obvious fact that ancient people also studied and got interested in the geometric properties of light and the portions of stereopsis. In sixteenth century, Leonardo da Vinci studied aspects of optics including eye anatomy and light reflection, and he even attempted to explain about stereoscopic depth perception. In "Trattato della Pittura di Leonardo da Vinci (Treatise of Painting)", which is published in 1584, mentioned that a given object occludes different parts of the background when they are observed by the left eye as compared to the right eye [4]. Eventually, in 1838, British scientist Sir Charles Wheatstone invented the first instrument designed to watch such images and produce a 3D effect. This device was to be called the "Reflecting Mirror Stereoscope". Wheatstone's stereoscope uses two centered mirrors at 45 degrees from observer's each eye, and they reflect each image to the eyes [5]. Interestingly, his device operates with a pair of drawing. This is because the photography was not yet available at that epoch. Anyway, his achievements inspire many photographic companies to get great interest in a new commercial market such as stereo photography and stereoscope.

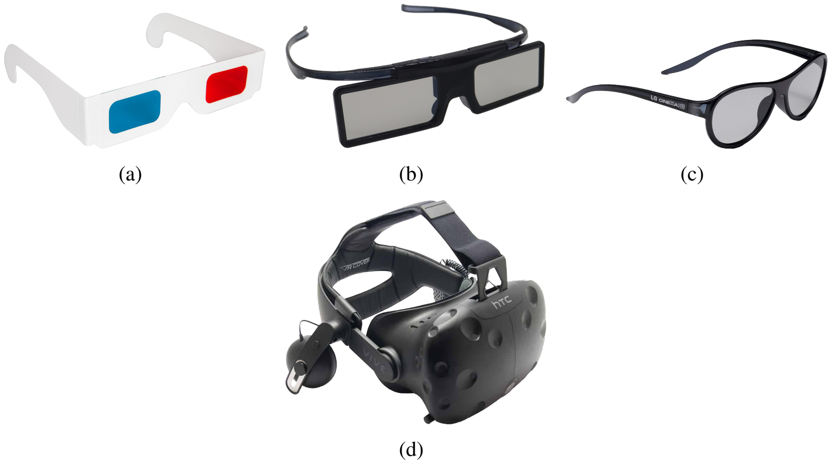
During the twentieth century, stereo photography became more popular and commercialized. A number of companies emerged who specialized in producing stereo images and providing 3D effect to viewers. Unfortunately, technology was not yet to the point of providing a good 3D experience without using glasses or utilizing an additional device for binocular observer. Anyhow, anaglyph is a widely used and applied approach in 3D imaging technique. This approach (as shown in Figure 1.2(a)) was developed in 1853 by W. Rollmann in Germany. Anaglyph images contain two differently filtered colored images red-cyan, green-magenta, or blue-yellow colors, in which these images have complementary contrast. When sighted through the color-coded anaglyph glasses, each of the two images are perceived by both eyes separately and the brain assembles both images. Finally, the pair of disparate images is integrated to a stereoscopic image [6, 7].

Another well-known approach for stereoscopic 3D system is to make use of passive or active 3D glasses. The passive glasses (a.k.a. polarized 3D system), as shown in Figure 1.2(c), use polarizing filters to create an illusion of 3D images by restricting the scene that fits with their respective information to each eye distinctly. These approaches cause each individual eye to see only the image made for that perspective, and vice versa [8, 9]. On the other hand, the active glasses (a.k.a. active shutter 3D system), which are shown in Figure 1.2(b), present the image intended for the left eye while blocking the right eye's view, and then presents the right eye's image while blocking the left eye's sight. This process repeats very quickly and sequentially (proper performance is over 30 images pair cycles per second with 60 Hz display system) so that the interruptions do not interfere with the perceived fusion of the two images into a single 3D image [10, 11, 12].

Each type of glasses has pros and cons. The active glasses provide the full image resolution to each eye entirely, and deliver completely separated scenes precisely. In fact, the polarizing filter of passive glasses cannot block out each classified images correctly in some specific situations. For instance, if there is some dark object beside something bright one in the displayed scene, an eye might notice the light which is intended for the other eye, thus, some interfered areas are appeared. The active glasses solve such 3D crosstalk or ghosting effect (like two superposed images) properly. On the other hand, the passive glasses are inexpensive and no batteries are required, so the weight is light. Besides, the passive glasses do not have a flickering effect, so that these types of glasses provide lesser dizziness. These active and passive types of glasses are broadly used in our real life nowadays because of their merits and advantages [13, 14].

On the other hand, during the last decades, head-mounted display (HMD) has also received big demand and attracted interest for the security and defense, military training, and even entertainment for the consumer's 3D experience. This device is mounted on the head or as part of a helmet, which has a small display to provide the scene for one or both eyes directly (see Figure 1.2(d)) [15, 16].

However, even though this high influential of the glasses-type 3D display



(a) (b) (c)

(d)

Figure 1.2: Various stereoscopic imaging devices: (a) Anaglyph glasses; (b) active 3D glasses from Samsung; (c) passive 3D glasses from LG; and (d) head-mounted display from HTC Vive.

system and HMD device, they have critical defects and limitations. First of all, glasses-type 3D system is only able to provide the uniformed 3D effect to the observers even when they are placed in various positions from the displayed scene. There is no degree of freedom to see the different perspective views, and only provides a fixed viewpoint. Therefore, the displayed scene seems to be too artificial and to be far apart from the real experience. Of course, recent HMD devices mount the gyro sensor and react the observer's head motion, but they are still clumsy to cover the human sense. Another critical defect is the visual fatigue. Even though there is a big advance of technology and big effort to solve this issue during recent days, the highlighted visual and motion sickness symptoms are not solved yet and many people are still suffering from this problem because of the sense of artificiality. The last flaw of these systems is that the observers obligatorily put on these equipments to their heads in order to watch the displayed 3D scene. Especially, HMD devices have lack of practicality because of its bulky size. These are inescapable facts and critical defects of early mentioned devices.

Meanwhile, there is another approach to see the 3D scene without needing any additional tool or wearable equipment. This is commonly known as autostereoscopy, which allows watching the 3D scene with the naked eyes directly. The main concept of this technique is that it provides numerous perspective views from the display, within a given range. Moreover, multiple observers can see the stereoscopic 3D scene with their own eyes, but each having distinct perspectives. Among others, new and interesting photographic method which is proposed first by Gabriel Lippmann in 1908 was remarkable. He presented the possibility of capturing and reconstructing the 3D scene by using an array of spherical diopters. He put this diopter array in front of the photographic film to register an array of distinct small images, each with different perspective information, and restored the captured scene in reverse order of pick up stage. He named this technique “photographie intégrale (Integral Photography: IP)” [17, 18]. After presented this remarkable research by Lippmann, great interests were awakened in many scientific groups and became a catalyst for research to autostereoscopic imaging. Since then, his research followers improved and complemented his achievements steadily, and some authors renamed it in different ways, such as integral-imaging (InI) [19, 20], light field imaging [21], or plenoptic imaging [22, 23].

In the past two decades, thanks to the advance of technology, some proposals of capturing and transmitting images in real-time were remarkable, and eventually, several companies announced their plenoptic camera (or light field camera), which are based on Lippmann’s IP theory. These cameras are able to capture the special image and perform various image modification functions provided by the manufacturers. Thanks to such new concept of photography techniques, these cameras have been spotlighted by many photographers and consumers, and even many scientific researchers also had a great interest. Above all, these cameras are able to compose a depth map based on the captured image. In fact, there are various types of depth-sensing cameras that are already published and broadly used in many different areas, but most of these 3D cameras mainly exploited additional technologies to extract the depth information. Besides, the depth map is one of the most important elements to describe and render our real world into the virtual 3D space. It is no exaggeration to say that the depth map is one of the core elements of the computer science in recent days. For that reason, the plenoptic camera has a great impact for many people because of such accessibility of use, performance, and even the novelty of technology.

However, even though there are a lot of great depth-sensing techniques and big advance of digital technologies, the conventional display system (as television, monitor, mobile phone, or even tablet PC) cannot provide the real 3D scene as the original volumetric information. This is because these display systems are only able to display the transmitted two-dimensional (2D) scene and that is their main objective. Here, our research was started from this aspect of the conventional display system. A lot of researches and approaches are already performed to compute and display the real world’s 3D information into digital devices. But at last, the final output was converted to 2D information and that's why the conventional display systems only provide the 2D scene. Therefore, we want to reproduce and display the real 3D information as it is, and provide the original 3D scene to the observers as closely as possible using the combination of computerized 3D information and InI technique.

The main aim of our proposal is to let the multiple observers see the restored real 3D scene with dynamic parallaxes, without any additional wearable equipment or glasses. This thesis addresses the challenges of well computerized 3D information based integral image production technique and algorithms, and concentrates in particular on the composition of a full-parallax 3D display system. On the other hand, we devoted the best endeavors to research to overcome the limitations of conventional InI technique. Based on the proposed experimentation, several depth-holes (or depthless pixels) refinement methodologies and recovering approaches are discussed and suitable solutions are also presented. Furthermore, we illustrate the procedures with some imaging experiments to prove the advantages of our approaches.

This dissertation is organized into six chapters. In Chapter 1, we start with a brief explanation about 3D imaging, and then narrate our research objectives and motivation. Chapter 2 gives a theoretical background of main components from this thesis. In Chapter 3, we provide several methodologies on how to compose an integral image from the computerized 3D data, in a more efficient and accurate way than in standard methods currently used in InI. Chapter 4 introduces our various approaches and algorithms on how to compute and compose the dense 3D data. Chapter 5 collects our main academic publications which build the basis for the research in this thesis. Finally, Chapter 6 concludes this thesis with a summary of the presented works, followed by a discussion of the future works.

Chapter 2

State of the art

2.1 Integral-imaging and pseudoscopic problem

InI is a promising 3D display technology due to its merits: it delivers continuous viewing points, full-parallax, and presents full-colored views to multiple observers in 3D space. In fact, the operating principle of InI is based on IP technique. One century ago, as we mentioned in Chapter 1, Lippmann's IP presented the possibility of recording the natural scene in 3D and displaying the restored scene using an array of diopters. Doing so, the developed photographic film is integrated floating in front of the diopters, and produced a 3D reconstruction of the original captured scene [17, 18].

On the other hand, in spite of this great merit of 3D scene restoration technique, it is important to remark that the composed image captured through IP technique is not directly projectable. The major issue for the 3D scene reconstruction based on IP is pseudoscopic (PS) problem, which is coming from the different directions and sequences of the pick up and display process, as shown in Figure 2.1. In pick up phase, the light rays scattered from the 3D object are passing through each lens, and compose elemental images (EIs). In display phase, the 3D object is reconstructed by the EIs with the same lens array as the one used in pick up phase. Interestingly, from the displayed scene, the observers will watch the rear side of the reconstructed 3D image which is in the direction from the 3D object to the lens array. This situation means that the final displayed 3D scene has 180°reversed depth. The main reason of this symptom is that a captured 3D object close to the lens array reconstructs the 3D scene closer to the displaying lenses, and further distanced object reconstructs the scene in further away from the lens array, respec-

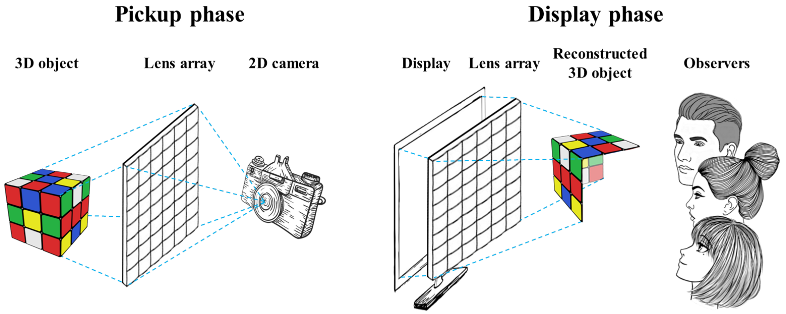
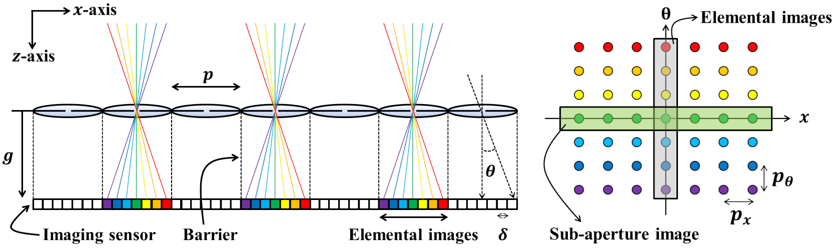


Figure 2.1: Scheme of IP technique to pick up and display. Through this direct pick up process, a reconstructed 3D scene has reversed depth. It is known as pseudoscopic problem. See text for further details.

tively. Therefore, the typical EIs without any additional process or transformation method cannot avoid PS problem. To overcome this fundamental drawback, various methods for the conversion from PS to orthoscopic (OR) of EIs (a.k.a. the PO conversion), have been proposed by many research groups [24, 25, 26, 27].

The main merit of InI is to transcribe the spatial and angular information of the rays that are proceeding from the 3D scene at the same time. Each ray of light can be described in a given point of its trajectory by the spatial coordinate $\delta_r = (x,y,z)$ of such point, and the angular inclination $\theta_r = (\theta,\phi)$ of its trajectory. Once we define the position and the inclination of the rays proceeding from a 3D scene when they pass through a given plane, they can easily be described through the radiance map (we sometimes name the radiance map as the plenoptic map). This fact permits us to reduce one dimension (in particular, a spatial coordinate) in the description of the plenoptic field. If we consider z as the direction perpendicular to the lens array, then we can represent the radiance map through the four-dimensional (4D) plenoptic function $R(\delta_r,\theta_r) = (x,y;\theta,\phi)$ [28, 29].

We illustrated this scenario in more understandable way via Figure 2.2. Based on Lippmann's scheme, the 3D scene is captured by the diopters array (a small lens array nowadays), like in Figure 2.2(a). For simplification, we assume that an array of pinholes substitutes the lens array, and we only consider the rays passing through each lens center with different incidence angles. This lens array is placed at a given distance g from the imaging sensor, and each lens of the array captures a 2D image of the 3D scene, but contains different perspectives, like as distinct colored rays in Figure 2.2(a). This group of individual perspective information be-



(a)

(b)

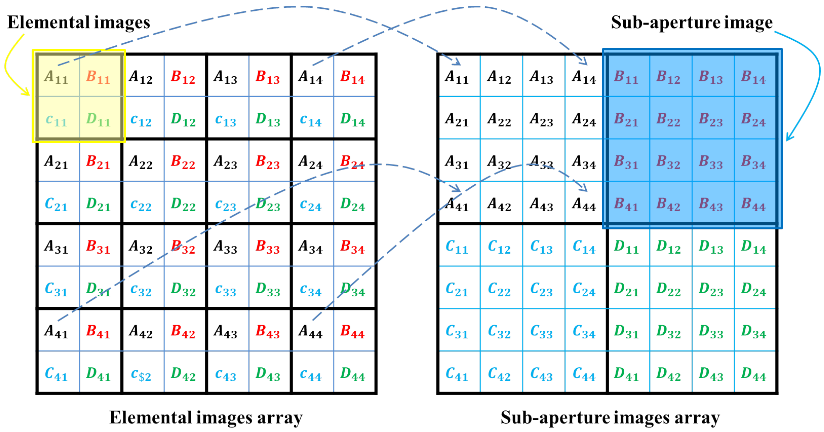
Figure 2.2: Scheme of integral photography: (a) Conventional InI system; and (b) corresponding sampled radiance map.

comes the set of the EIs, and hereafter, the whole collection of EIs will be referred to as the integral image. Thanks to the advance of technology and the big interest about InI, the diopters array is substituted nowadays by a micro hemispherical lens array (a.k.a. Microlens array: MLA), and hereafter, the captured EIs are also referred to as the microimages in further text. In order to avoid overlapping between the different EIs, a set of physical barriers is required.

On the other hand, this captured integral image can be represented in a graphic scheme, like as in Figure 2.2(b). This sampled radiance map is determined by the gap g, the lens pitch p, and each pixel size of the imaging sensor δ_p. The sampling period in the spatial direction is given directly by p, and the angular one p_θ is given by the following simple equation:

$$p_\theta = \delta_p / g. \tag{2.1}$$

From this radiance map, we can find the EIs from each column of the sampled field. On the contrary, each row of the sampled field corresponds to not only a set of rays passing through the lenses with same incidence angle, but also located in an equidistant position in the imaging sensor. Thus, any horizontal line in Figure 2.2(b) can be grouped to form a sub-aperture image (or sub-image) of the 3D scene. Interestingly, every sub-aperture images contain the orthographic (or orthogonal) views of the 3D scene. For these views, there is no perspective distortion wherever the viewpoint is placed, and all sub-aperture images have same scale with an identical field of view (FOV) of the scene, while, all of sub-aperture images have different perspectives. Thus, we could say that the sub-aperture images are equivalent to those obtained with an array of telecentric cameras whose optical axis are inclined accordingly.



A_{11}	B_{11}	A_{12}	B_{12}	A_{13}	B_{13}	A_{14}	B_{14}
C_{11}	D_{11}	C_{12}	D_{12}	C_{13}	D_{13}	C_{14}	D_{14}
A_{21}	B_{21}	A_{22}	B_{22}	A_{23}	B_{23}	A_{24}	B_{24}
C_{21}	D_{21}	C_{22}	D_{22}	C_{23}	D_{23}	C_{24}	D_{24}
A_{31}	B_{31}	A_{32}	B_{32}	A_{33}	B_{33}	A_{34}	B_{34}
C_{31}	D_{31}	C_{32}	D_{32}	C_{33}	D_{33}	C_{34}	D_{34}
A_{41}	B_{41}	A_{42}	B_{42}	A_{43}	B_{43}	A_{44}	B_{44}
C_{41}	D_{41}	$C_{\$2}$	D_{42}	C_{43}	D_{43}	C_{44}	D_{44}

A_{11}	A_{12}	A_{13}	A_{14}	B_{11}	B_{12}	B_{13}	B_{14}
A_{21}	A_{22}	A_{23}	A_{24}	B_{21}	B_{22}	B_{23}	B_{24}
A_{31}	A_{32}	A_{33}	A_{34}	B_{31}	B_{32}	B_{33}	B_{34}
A_{41}	A_{42}	A_{43}	A_{44}	B_{41}	B_{42}	B_{43}	B_{44}
C_{11}	C_{12}	C_{13}	C_{14}	D_{11}	D_{12}	D_{13}	D_{14}
C_{21}	C_{22}	C_{23}	C_{24}	D_{21}	D_{22}	D_{23}	D_{24}
C_{31}	C_{32}	C_{33}	C_{34}	D_{31}	D_{32}	D_{33}	D_{34}
C_{41}	C_{42}	C_{43}	C_{44}	D_{41}	D_{42}	D_{43}	D_{44}

Figure 2.3: Elemental images array and sub-aperture images array are mutually convertible each other.

On the other hand, we can confirm that the number of pixels in EIs is identical to the number of viewpoints from each lens, and the total number of EIs in the integral image is equal to each sub-aperture image's resolution. Here, it is remarkable that the collection of EIs and sub-aperture images are mutually convertible. The way of conversion between the integral image and sub-aperture images array is very simple. As we illustrated in Figure 2.3, a simple image processing procedure for pixel rearrangement is required. For instance, the collection of left-top pixel (A_{ij}) from every EIs are transposed and assembled to the first sub-aperture image, and every right-top pixel (B_{ij}) become an element of second sub-aperture image, respectively. Furthermore, the full image size of both, the integral image and sub-aperture images array are equal. Figure 2.4 provides this conversion procedure with an experimental result and details.

The sub-aperture images array is exploited in many different research fields. Among them, research approaches are performed in image refocusing, image quality and resolution enhancement, occlusion removal and restoration, object tracking in heavily occluded situation, as well as depth map calculation from sub-aperture images, 3D travelling movie composition, etc [25, 26, 30, 31, 32, 33, 34, 35, 36]. We also exploited the sub-aperture images array in our experiments. We will narrate our approaches by using this sub-aperture images array in further chapters in detail.

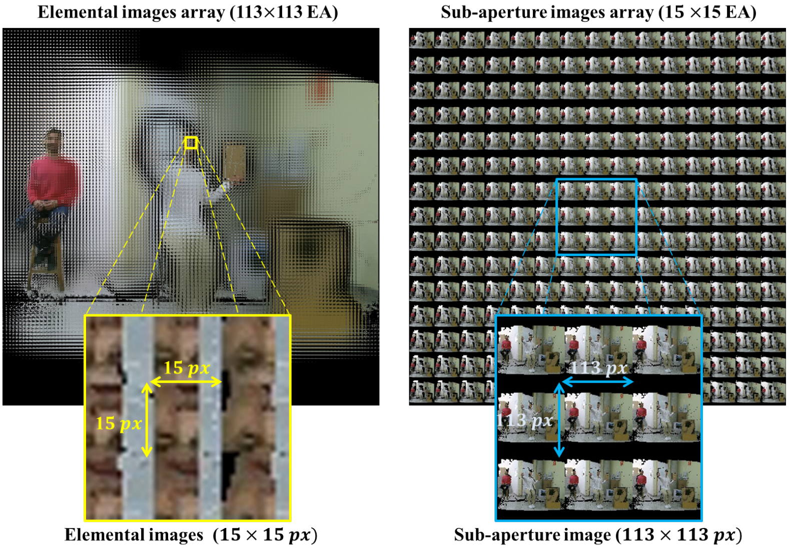

Elemental images array (113×113 EA)
Sub-aperture images array (15 ×15 EA)
15 px
15 px
113 px
113 px
Elemental images (15 × 15 px)
Sub-aperture image (113 × 113 px)

Figure 2.4: Conversion example between the elemental images array and the sub-aperture images array. Left integral image contains 113 × 113 elemental images (with 15 × 15 pixels each), and the right array contains 15 × 15 sub-aperture images (with 113 × 113 pixels each). But, both images have same size (1695 × 1695 pixels).

2.2 3D cameras and plenoptic camera

A methodology on how to record and store the natural scene in 3D as being itself is one of the most attracted and spotlighted research task during last decades. This phenomenon was a natural part of the process in the development of digital age, and it also big endeavors were needed to extend the 2D photography further by a lot of great pioneers. Among all, a highlighted technique is the stereo vision, which takes advantage of the disparity information from two aligned cameras. In fact, the stereo vision technology is based on a similar concept as binocular observer's physical formation, and also resembles a biological process of stereopsis at the depth estimation step. This camera configuration has been the representative of depth imaging technique for a long period, and various related researches are still discussed and studied today in order to improve the accuracy of depth estimation result [37, 38, 39, 40].

This approach encodes the difference in horizontal (or vertical) pixel coordinates of corresponding image points at the image pair. It means that every calculated disparity values are in pixel unit, and all of extracted disparity values compose a group of disparty information. These collected information is translated to a disparity map, and the values in this map are inversely proportional to the depth at the corresponding pixel location. FLIR Systems company explains the method on how to determine the depth of a pixel based on the computed disparity value in [41]. The formula is the following:

$$Z = fB/d \tag{2.2}$$

where Z is the depth distance along the camera in z-axis (in metres), f is a focal length (in pixels), B is the baseline that is the distance between the two aligned cameras (in metres), and d is the disparity value (in pixels). For instance, if the disparity value is big, the calculated depth distance is short, and vice versa. This fact certifies that the disparity value is inversely proportional to the depth distance. Furthermore, when Z is calculated, the rest of real 3D coordinates, X and Y, also can be derived via usual projective camera equations, as following formulas:

$$X = uZ/f \tag{2.3a}$$

$$Y = vZ/f \tag{2.3b}$$

where u and v are the relative pixel location at the 2D image, and (X,Y,Z) are computed in real 3D coordinates along the camera's position. Here, note that u and v do not share the same coordinate system as target pixels' coordinates of the 2D image. The (u,v) pair is calculated by the center pixel position of width and height of the 2D image, with following formulas:

$$u = img\_xPos - img\_WidthCenter \tag{2.4a}$$

$$v = img\_yPos - img\_HeightCenter \tag{2.4b}$$

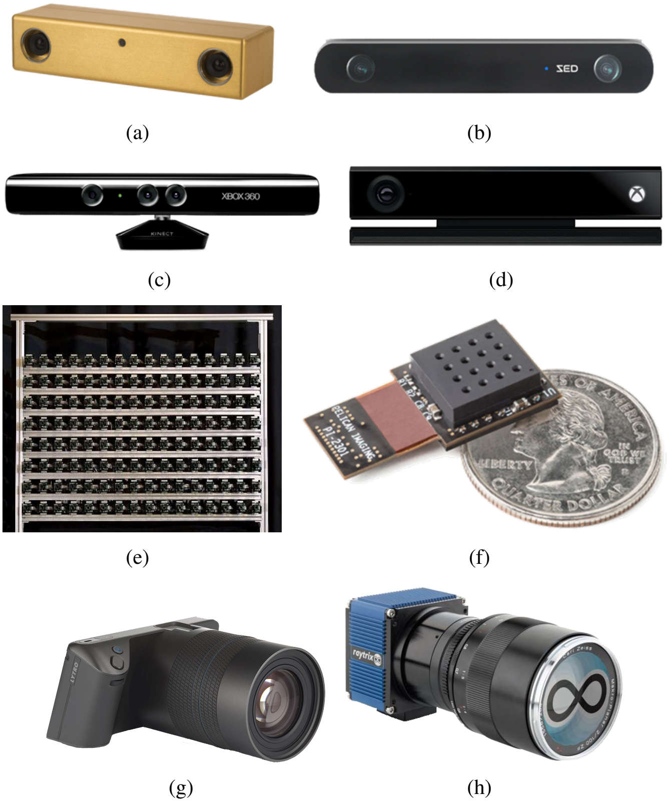
where img_xPos and img_yPos are each target pixel coordinates of the 2D image, $img\_WidthCenter$ and $img\_HeightCenter$ are 2D image's center pixel position of width and height. So in sum, $(u,v), f$ and d are all in the pixel unit, and $(X,Y,Z), B$ are all in metres, respectively.

However, in order to calculate the disparity map via the stereo camera configuration, people have to calibrate the cameras, derive the fundamental matrix, extract the camera's inherent parameters (intrinsic and extrinsic parameters), and rectify the source images, etc [42, 43, 44]. These procedures are complicate and the experimental system is not easy to stabilize. For instance, if the camera tilts or even shifts its position a little after already performing the camera calibration, one has to reprocess the whole step from all over again. Fortunately, several companies launched their own stereo camera system and simplified the modification procedures in order to manipulate the stereo cameras in a much easier way, although the price is quite expensive and not affordable (see Figures 2.5(a) and (b)) [37, 38]. For that reason, it was not easy to manipulate these commercialized stereo cameras for many small research groups and young researchers, and much less students for their academic usage.

Incidentally, the depth-sensing technologies related to the infrared (IR) light sensor are spotlighted during last decades. The IR depth-sensing camera allows to acquire the depth information in real-time with a high frame rate. This IR sensing type depth camera system has caught the attention of many people due to its accessibility and easier usability, as compared with the stereo camera system. In spite of these merits, however, the IR depth-sensing camera was also very expensive at the beginning. Recently, after being launched and commercialized by several big companies, it is easy to find and purchase the various brands of IR depth-sensing cameras in the commercial market with affordable prices [45, 46, 47, 48, 49, 50]. Thanks to these advanced low-cost cameras, people manipulate these cameras into their studies and researches, and eventually, a great interest of 3D imaging technology has arrived.

Among others, Kinect sensor, which is developed by Microsoft and Israel Company PrimeSense, got really big attention. In fact, in 2011, Kinect was sold 8 million units in its first 60 days on the market, and claimed the Guinness World Record of being the "fastest selling consumer electronics device" [51]. Two versions of Kinect devices are released up to now: Kinect v1 (Kv1) in 2010 (Figure 2.5(c)), and Kinect v2 (Kv2) in 2014 (Figure 2.5(d)). Interestingly, the original purposes of these cameras were detecting the human body's gesture, acquiring the voice command, and recognizing the user's facial expression [45, 46, 52, 53]. It means that these cameras enable the users to control and interact with the entertainments, using their body movements and spatial gestures, instead of the necessity of the additional controllers or joysticks. However, many people manipulate these cameras into their researches because of the depth acquisition performance and accuracy.

Kv1 and Kv2 have completely different measurement principle for obtaining a dense depth map. In detail, Kv1 uses a structured IR light pattern emitter, and then the IR camera captures and calculates the depth distance through the acquired pattern information. The depth measurement is performed by a triangulation process, and the distance between picked speckles patterns at the captured scene become the disparity information, as described in [54, 55]. In comparison, Kv2 utilizes



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 2.5: Various type of 3D cameras: (a) Bumblebee stereo camera from FLIR Systems; (b) ZED stereo camera from Stereolabs; (c) and (d) Kinect v1 and v2 from Microsoft; (e) multi-camera array; (f) PiCam camera from Pelican Imaging; (g) Lytro plenoptic camera from Lytro; and (h) Raytrix plenoptic camera from Raytrix. Most of them are already commercialzed and easy to purchase via the internet, except (e) and (f). (e) is configured by Stanford university, and (f) is no longer available to purchase. See text for further details.

the Time-of-Flight (ToF) technology, which exploits emitting IR beams pulsed at high frequency rates. From the reflected IR light from the most 3D surfaces, the sensor evaluates the depth distance by measuring the IR flash's returning duration [56]. Of course, both, the RGB camera and IR camera of Kv1 and Kv2 also have different image size and FOV information. Even though several specifications are released from the manufacturer, further detailed information are not provided to the consumers, like the coupled areas of the scene between the RGB and IR cameras, FOV of the RGB camera, etc. To confirm and analyze these issues, we performed a simple experimental comparison between Kv1 and Kv2 with several referenced information [52, 53]. Figure 2.6 shows the difference of FOVs between the RGB and IR cameras of Kv1 and Kv2. We defined a common optical axis for both devices, and set both Kinects in parallel from a common target, and placed in the same position. We choose a chessboard pattern as a common target, which has simple and repetitive shapes, and permits to detect the feature points easily. Most of all, the regularized pattern influences to improve the accuracy of calibration's result [57]. We find the common correspondence feature points in each captured image and calculate the correlation parameters that define the so-called homography matrix (projectivity or projective transformation). These parameters represent a general plane-to-plane correlation equation in a projective plane [58, 59]. Finally, these derived values allow to map from 2D view of one camera to another, and to confirm the coupled areas of the scene based on the FOVs of Kv1 and Kv2's cameras.

In despite of their advantages, these IR depth-sensing cameras have some drawbacks and limitations. First of all, these cameras do not work properly in outdoor environment. The main reason is that the outdoor ambient light (like sun light) also contains IR light. It means that there are a lot of interferences between the different sourced IR lights, and thus, a captured depth map will not have enough good quality and dense information, as compared to the original capability. Another limitation is that these IR depth-sensing cameras have their limited depth volume capacities. The way of depth acquisition method in every IR depth-sensing devices are different, but, commonly these cameras have a maximum acceptability and limitation due to their inherent specifications. The last flaw of these cameras system is that they shall be connected to the electrical supply equipment, so that they are not conveniently portable.

In the meantime, as we mentioned in Chapter 1, several research groups and companies have released their special cameras which applied Lippmann's IP theory [60, 61, 62, 63, 64]. Among them, PiCam from Pelican Imaging (Figure 2.5(f)) received big attention because of its very tiny modularized cameras array. The PiCam provides 16 different perspective images with synthesized 8 Mega pixels RGB image, and even a depth map which is estimated by using various image enhancement algorithms and synthesization skills, as described in [61, 62]. However, in spite of this camera module's great performance and competitive price, it is no longer available to purchase any more. On the other hand, Lytro camera from Lytro

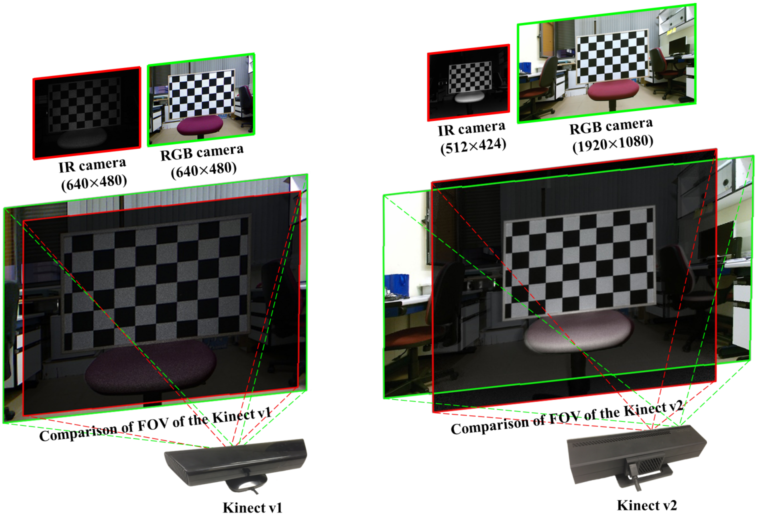
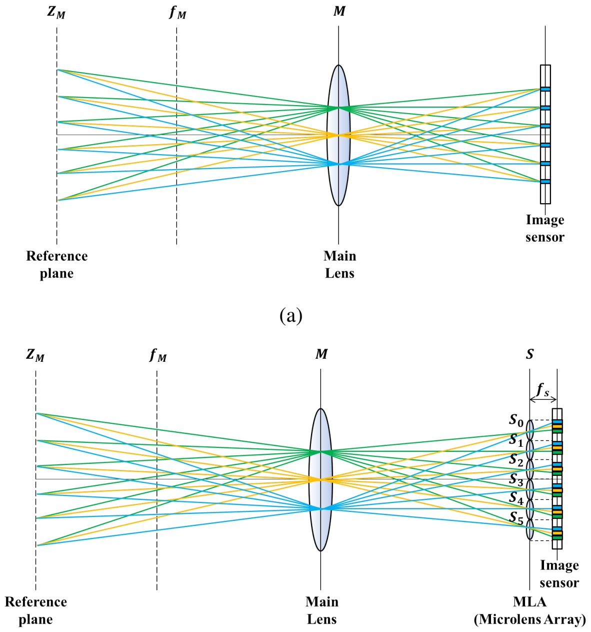


Figure 2.6: Kv1 and Kv2's overlapped region at the captured scene. In green rectangle is represented by the RGB camera's view, and red rectangle is the IR camera's view, respectively. Both, Kv1 and Kv2 have different FOV and image size, and the depth acquisition method is also different.

(Figure 2.5(g)) and Raytrix camera from Raytrix (Figure 2.5(h)) are regarded as the best-known commercialized plenoptic cameras in the world. These cameras also provide a dense depth map from a single shot captured plenoptic image via their veiled technologies.

In fact, we would say that there are two methods of capturing a plenoptic field based on Lippmann's photography scheme. One is a synchronized multi camera array and the other is to perform a simple modification of a single camera. The important feature from these two methods is that they are not demanding any additional special lighting emitters or sensors, and they just pick up the 3D scene like as normal digital cameras contrary to the IR depth-sensing cameras. The first method has an advantage of allowing the capture with big parallax. However, multi camera array system has several dificulties: this composition becomes bulky like as Figure 2.5(e), and not only requires to synchronize all cameras, but also the management of the huge amount of acquired data. As an alternative way, the second method is very useful when small parallaxes are acceptable. This system works only inserting an array of microlenses in the image plane, and shifting the imaging sensor axi-



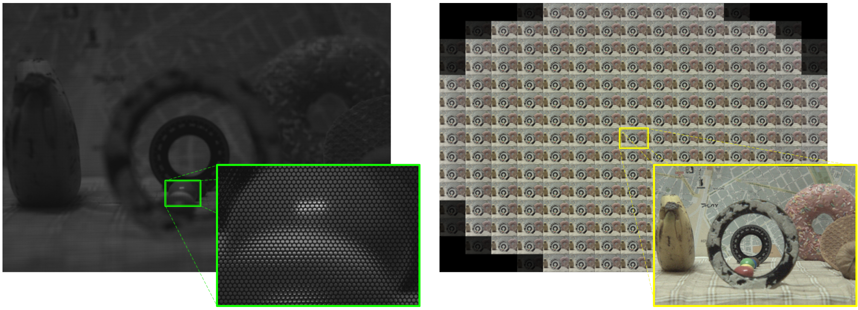
(a)

(b)

Figure 2.7: Scheme of image capturing system with two cameras: (a) Conventional camera; and (b) plenoptic camera. In fact, (a) contains the angular information, but discarded because of the pixel overlapping problem in same position at the imaging sensor. On the contrary, (b) can record both, the spatial and angular information thanks to the insertion of the microlens array.

ally. Therefore, this configuration does not need any synchronization procedure at the capturing scheme. This camera composition is called the plenoptic camera (or light field camera), in which Lytro and Ratrix cameras examples in this category [60].

Figure 2.7 shows the difference of capturing scenario between the conventional camera and plenoptic camera. In fact, this is little bit far from the real situation of the general optical system, in which the objectives may be composed by the combination of numbers of converging and diverging lenses, built with different glasses that having various focal lengths, and occurred a large aperture stop to consider the various incidence angles of light. However, in spite of such different



(a)

(b)

Figure 2.8: (a) Raw plenoptic image captured by the plenoptic camera (Lytro Illum); and (b) composed sub-aperture images array from the captured plenoptic image. We can see that the captured raw plenoptic image from Lytro camera (a) forms in a hexagonal shape. In fact, the hexagonal grids of MLA assist to capture a higher number of microimages from the scene.

conditions, all these optical elements can be substituted by a single thin lens with same cardinal parameters, namely, reference plane, focal points, and *f-number*. The main difference between Figures 2.7(a) and (b) are the insertion of a MLA at the image plane.

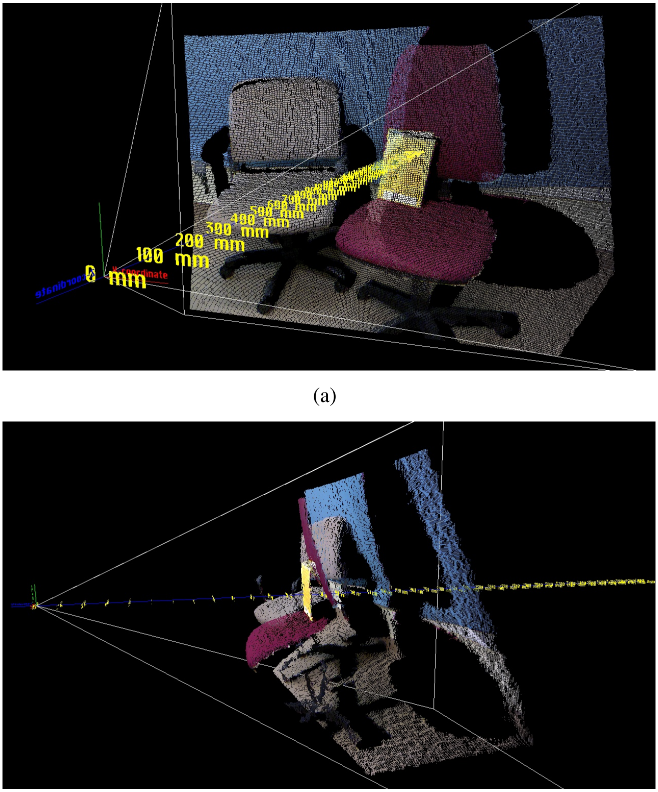
Note that the plenoptic image captured by the plenoptic camera is also convertible to sub-aperture images array. As we mentioned in previous section, the sub-aperture images array is used in many different research areas, and it had very practical applications for our researches during past years, too. In our experiments, we exploited Lytro camera and proceeded several research tasks. However, we faced some critical problems. In fact, the captured raw plenoptic image using Lytro camera contains a grayscale image, and the microimages are arranged in a hexagonal shape (see Figure 2.8(a)). The main difficulty when dealing with this camera is that there is a Bayer color filter array over the camera's sensor to capture the intensity of light with different color spectrum, that follows the conventional rectangular geometry. Thus, firstly, it must be demosaiced to get the color information back, and then composed the coloured sub-aperture images array [65]. To perform these procedures, we mainly adopted the algorithms in [66, 67] and the provided toolbox. Figure 2.8(b) shows our experimental result. In further chapters, we will narrate the methodology on how to perform the depth map estimation by using this sub-aperture images array.

2.3 Point cloud

As indicated above, the conventional camera transcribes and saves the captured natural scene in 3D into the plane 2D information like as a paper or a thin sheet. Thus, it cannot avoid losing the original volumetric 3D information. Moreover, the simulated 2D information cannot present and interpret the occluded or concealed areas properly, and even, it is unable to observe the veiled parts in different perspective views either. On the contrary, the computerized 3D data contain the volumetric information and overcomes the early mentioned limitations precisely. The most commonly used methodology to compose and visualize the 3D data is called the point cloud approach [68, 69]. Basically, each point from a point cloud contains its coordinate information (x, y, z) with the color intensity (R, G, B), and a set of points form a 3D shape in the space. Thanks to the composition of the point, it is convenient to simulate and handle the composed point cloud in the computer-generated virtual 3D environment. Various research areas are already performed and applied to the point cloud into their research tasks, such as, geometry, architecture restoration, virtual reality, manufacturing, 3D printing, robotics, biomedical engineering, computer vision, etc. Thus, it is no wonder that the point cloud is one of the essential elements in present day to describe and render our real world into the virtual 3D space.

On the other hand, a point cloud can be extended via fusion between two, or even more sets of point clouds using a point set registration method. In fact, the registration methodology calculates and helps to fuse between paired 2D images (or 2D shapes) or a number of point clouds, so that merged information become denser and vacant areas of point cloud are recovered by complementing each other. Due to such merits, this method is still treating as one of the most active research tasks in the world [70, 71, 72].

We also exploited a point cloud set into our experiments. We composed a point cloud set by using the combination between the captured depth map and its correspond color image, and generate an integral image via InI concept in the virtual 3D space. We will narrate our methodology on how to produce the integral image from the composed point cloud, and also provide the configuration of our proposed integral-imaging display system in further chapters.



(a)

(b)

Figure 2.9: Display of a set of point cloud in the virtual 3D space. From both subfigures, it is clear that each point has its own (x,y,z) coordinates with the color intensity (R,G,B). All of points are arranged along the camera's position. Note that the displayed point cloud is captured by Kv1.

Chapter 3

Integral image production from point cloud and display

Our main goal in this chapter is to compose an integral image based on a set of point clouds, and provide a good quality 3D scene to multiple observers by using our proposed integral-imaging monitor. In fact, our methodology on how to compose the integral image is originated from IP technique, but we applied and extended the algorithm with several further particular approaches. In addition, we exploited a computer acceleration technique to solve the heavy computation time at the moment of integral image generation scheme. In the first section, we will narrate our methods on how to compose the integral image using a point cloud with differentiated approaches. The way of boosting the computation speed will be presented in the following section, and then we will provide the composition of our integral-imaging monitor with some details in the last section.

3.1 Production of integral images using point cloud

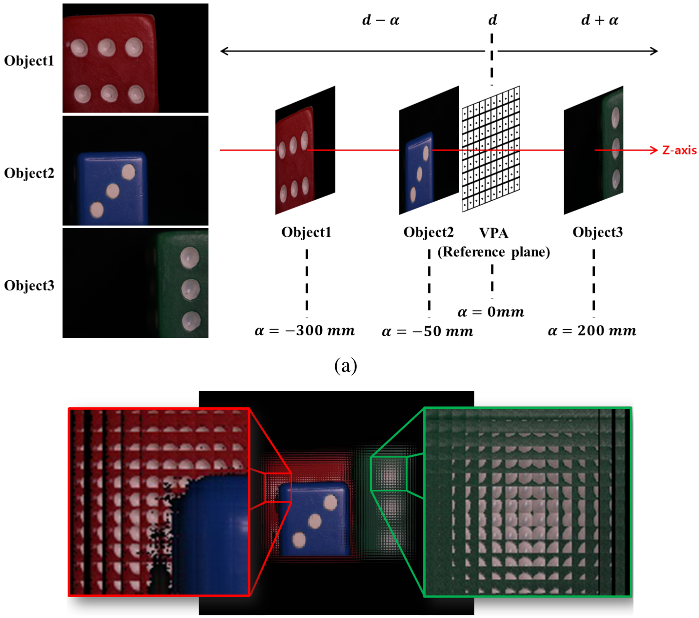
There have been many trials to make a good quality integral image since Lippmann's IP technology was announced. However, in spite of the advancement of science and technology, conventional InI has several drawbacks. Among them, experimental environment of InI is confined in the real world's hardware system. Besides, most components of the system are not flexible depending on the variable situations, even, the acquired information is not modifiable in the post-processing step. In this sense, we moved our research stage to the virtual 3D space, that is, experimental parameters are adjustable whenever and wherever we want. This novel approach was not actively discussed and studied last years, what confirms the novelty of our research.

Our proposal is started from the curiosity on how to compose and visualize the 3D data in efficient way. To reach this goal, we needed to solve several research issues, such as, the methodologies on how to manipulate and visualize the 3D information, how to compose an integral image in the virtual 3D space, and how to solve PS problem precisely.

We first compose a dense point cloud using various 3D cameras, and simulate them in the virtual 3D space in order to check the acquired 3D data and analyze their formation in a more intuitive way. We set our experimental system to compose a point cloud via the captured RGB image and depth map from 3D cameras in real-time, save and load the composed point cloud simply and handy to browse the simulated point cloud where we want to look at. After that, we formulate a methodology on how to compose an integral image from the simulated point cloud. We investigated various related researches to find a relevant technology, and we decided to adopt the smart pixel mapping method from [73], especially, the usage of the virtual pinhole array (VPA) concept at the synthetic capturing phase. The VPA can provide the synthetic information using its tunable parameters. This pinhole array resembles an array of cameras or a lens array, but the VPA has particular properties. The components of VPA are configurable, such as the FOV, the pitch, the gap between the position of pinhole and the virtual imaging sensor, and the location of VPA.

Afterwards, we need to perform the corresponding post-processing. The image captured by the pinhole (VPA) camera has reversed and inverted PS orientation, as opposed to the original formation and orientation. This phenomenon is well-known as the response of a ”camera obscura”. When an illuminated scene is projected through a small hole, the penetrated light forms a flipped scene (left to right, and upside down) [74]. To solve this distortion, we follow Okano [24]'s approach. We rotate each EI composed via the VPA by 180°about its center. Interestingly, this simple image processing procedure avoids this critical defect of IP tehcnique which is the well-known PS problem. By applying Okano's approach, the depth-reversed integral image composed by IP technique is turned-over again, and thus, the final refined integral image does not have any PS problem, and it has an accurate depth volume at the displayed scene.

To check this approach, we first set a single virtual pinhole camera along an imaginary plane in the virtual 3D space, and then pick up a scene in order to verify its composed result and check the feasibility of our algorithm. After confirming the performance of this virtual pinhole camera, we increase the number of cam-



(a)

(b)

Figure 3.1: Simple synthetic simulation to compose an integral image: (a) Brief scheme of the proposed experimental environment; and (b) composed integral image captured by the virtual pinhole array. Note that the position of pinholes is represented by the reference plane in this configuration. All processes are performed in the virtual 3D space. See text for further details.

eras and proceed a simple simulation, shown in Figure 3.1. We use 2D images as test target objects in 3 different positions along the *z-axis*'s direction, as shown in Figure 3.1(a). Next, we perform a projection procedure from each object's pixels through every virtual pinhole sequencially, from negative ($\alpha < 0$) to positive ($\alpha > 0$) direction. Figure 3.1(b) shows the projection result. In this figure, we can confirm that 3 different placed objects are composed by the EIs with distinctive conformations. In detail, the part of object 1 (red dice), which is projected firstly to the VPA between 3 objects, contains small portions of information from the original scene in their adjacent EIs. The area of object 2 (blue dice), which is positioned just in front of the VPA, contains its well-focused scene with clear shape. We could say that the object 2 penetrates itself entirely to the VPA without any

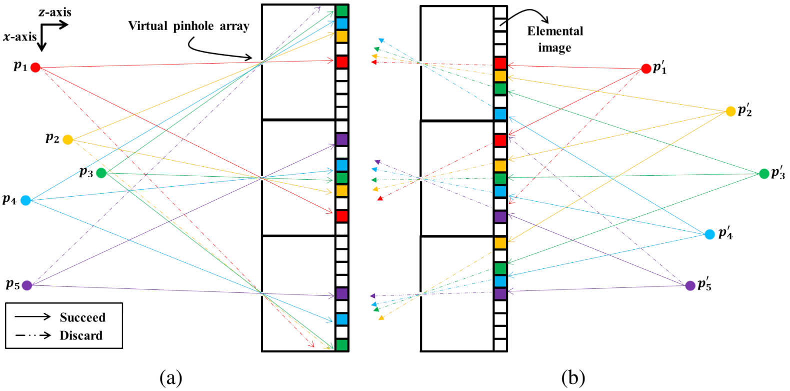
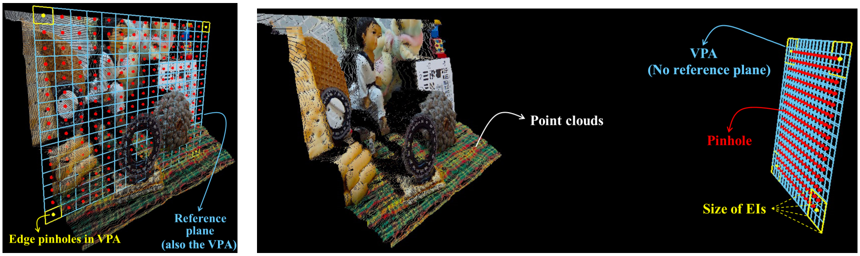


Figure 3.2: Comparison of the projection scheme between two different situations: (a) When points are placed in front of the VPA; and (b) when points are located behind of the pinholes. Both, (a) and (b)'s projection schemes are completely different. If light coming from a point does not reach to the imaging sensor or impact to the barrier first, the point will be discarded.

invasion of adjacent areas of the EIs. Thus, it is apparent that the pixel assignment composition is strongly dependent on the VPA's position. Finally, the part of object 3 (green dice), which is located after the VPA and projected to pinholes in the last sequence between 3 objects, is composed with inverted EIs. In fact, it is a very unusual situation in real world to capture the scene along the opposite direction that the camera is looking at. In virtual 3D space, on the contrary, it is possible to acquire the scene in such uncommon situation thanks to the evasion from the laws of physics and nature, and simplification of real world's complicate conditions. Therefore, we could perform such projection from the object 3's pixels to the VPA directly without any problem.

We can explain these different projection schemes by means of Figure 3.2 with some examples. Every point (p'_1 to p'_5) in Figure 3.2(b) is positioned behind of the VPA, a similar situation as object 3 in Figure 3.1(a). When the projection trajectory impacts first on the imaging sensor and then on the pinhole, a chosen pixel becomes a component of EIs. This sequence is a completely opposite situation to Figure 3.2(a). Every point (p_1 to p_5) in Figure 3.2(a) are located in front of the VPA (objects 1 and 2 in Figure 3.1(a) correspond to this case). These points penetrate to each pinhole first with different incident angles, and then impact the virtual imaging sensor later. Thus, Figures 3.2(a) and (b) present completely contrasting



(a)

(b)

Figure 3.3: Two different approaches to compose an integral image using a point cloud and the VPA: (a) VPA is placed inside of the point cloud; and (b) VPA is located far from the point cloud. The position of VPA in (a) decides the position of reference plane directly; however, this is not the case in (b). Red dots indicate the pinholes, yellow dots are edge positioned pinholes in the VPA, and yellow quadrangles show the size of composed image via the pinhole camera, respectively. The VPA in both, (a) and (b), have same number of pinholes with equidistant positions (15 × 15), but different gaps in the two approaches.

situations. After all projection procedures are done, the composed EIs are rotated 180°about its center position, following Okano's approach. For that reason, the object 1 and 2's composed EIs have the same orientation as the original scene, and object 3's EIs show an inverted formation, respectively. We can observe these results with some details when we display the computed integral image using an integral-imaging monitor. After displaying of Figure 3.1(b), the observation of either orthoscopic real (floating outside from the monitor) or virtual (inside of the monitor) images are strictly determined by the position of the VPA (d in Figure 3.1(a)). From now and hereafter, the plane where the VPA is placed will be regarded as the reference plane when the pinholes array is located inside of the scene under issue. Besides, the reference plane will be identical to the same position of an integral-imaging monitor.

We used two methodologies to compose an integral image using a point cloud and the VPA. The first method is to put the VPA inside of (or close to) the point cloud and perform the projection scheme from each point to all pinholes one after another (see Figure 3.3(a)). The second strategy is to dispose the VPA far from the point cloud and pick up the scenes (see Figure 3.3(b)), and then compute the integral image. To the best of our knowledge, each method has pros and cons with distinct differences.

The first approach composes the integral image with great parallaxes, and provides an abundant depth volume at the displayed scene. In fact, the first approach has farther gap between each pinholes than the second approach; thus, EIs have good disparities. Besides, the way of placing the VPA inside of the point cloud has the great merit to select a reference plane in a much simpler and more intuitive way. However, a problem with this method is that a point which is placed really close to the VPA (in front or behind the pinholes) cannot be projected to other pinholes because of the limited incidence angles. The neighboring points from the VPA also cannot be penetrated to their adjacent pinholes, and therefore, these points form some apparent black areas (see Figure 2.4's left image).

The second approach works similarly to the synthetic aperture method. In fact, that method utilizes only a single digital camera to capture all scenes, and the camera is mechanically displaced to acquire the different perspectives [75]. On the contrary, in our proposal, we can virtually acquire each different perspective images via the VPA directly, and then compose the integral image using the conversion process between the sub-aperture images and the elemental images array, as shown in Figure 2.3. This approach is able to produce the EIs without any black pixels, and it provides a large depth of field (DOF) at the displayed scene. However, this second approach also has a critical defect. If a gap between pinhole cameras is bigger, the DOF is reduced, and a computed integral image provides a blurred image at the displayed scene. On the contrary, if the gap is shorter, the parallaxes are mutually decreased (the displayed scene will be shown like a plane 2D image). That's why the second approach must be underwent many trials and errors in order to find not only the adequate parameters but also the reference plane and the VPA's proper position.

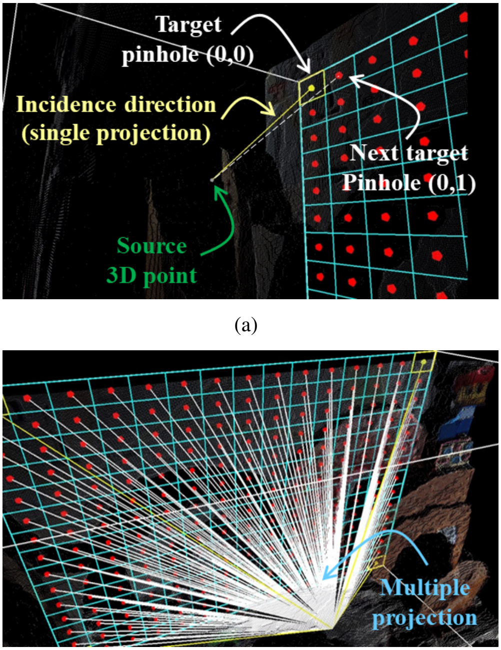
So in sum, the first approach is suitable to be applied when the visualized scene needs great parallaxes, abundant depth perception, and the necessity of tuning the reference plane's position frequently, rather than a large DOF of the scene. On the other hand, the second approach is appropriate to be exploited when the displayed 3D scene needs not only great image quality with uniformly-focused objects but also demands certainly defined reference planes, instead of the dynamic viewpoints and great 3D perception.

3.2 Boost the computation speed using graphics processing unit

One of most usual hardware acceleration techniques is the well-known graphics processing unit (GPU) accelerated computing. GPU acceleration technique is based on the use of computer hardware specially aimed to process some vast amounts of data quickly, used to perform heavy computations more efficiently than is possible in proceeding on the general-purpose central processing unit (CPU). Especially, GPU computing assists to provide an improved performance when used in operations adapted to the application-specific hardware designed system. Thus, GPU is used mainly in some of the compute-intensive and time consuming portions of the codes, and the rest of applications are run on CPU concurrently. In fact, CPU is in charge of organizing an entire operating system and managing various applications, contrary to GPU that is specialized-independent device to accomplish its particular missions. For that reason, CPU and GPU can be roughly split into data-parallelism and task-parallelism, in which data-parallelism is applying the same operation to multiple data-items, and task-parallelism is doing different operations simultaneously. So in sum, GPU is designed for data-parallelism, while CPU is designed for task-parallelism.

The use of GPU computing into the scientific applications and researches was started quite recently. At the beginning, graphics chips were designed for performing some fixed-functions in graphics pipelines. However, there are currently big demands to use GPU in general-purpose applications. Over the years, these graphics chips became increasingly programmable over the 90s, and many researchers and scientists from various areas started using GPU for their general-purpose computing. This was the advent of the movement called the general-purpose computing on GPU or GPGPU [76]. Several computer graphics chip companies realized and payed attention to the potential of bringing this performance to enlarge the scientific community, and they did great endeavor in modifying GPU to make it fully programmable for scientific purposes and applications. Thanks to those big efforts, we can use and apply GPU computing techniques into our researches much easier in recent days.

One of the great benefits of using a GPU is parallel computing. In some benchmarks result [77], GPU's parallelism performance has been shown to be 18 times faster than CPU. This enormous difference in performance comes from the distinct numbers of cores from CPU and GPU. Nowadays, recent CPUs have multi-core (2-dual, 3-triple, 4-quad, 6-hexa, 8-octa, 10-deca, and 12-dodeca, respectively) and the number of cores is still growing steadily. On the contrary, in spite of using many cores from CPU, the computation parallelism is not suitably compared to GPU. In fact, on CPU there are 1 or 2 threads per core and GPU has 4 to 10 (a thread is a kind of worker that performs its own missions, and completely independent from other threads). GPU has thousands cores, that means that GPU has more than 10000 active threads that are ready for running their tasks in parallel. Thus, literally, there are enormous differences in the number of threads used between CPU



(a)

(b)

Figure 3.4: Scheme of our approach on how to calculate the impact points as a function of the incidence angle: (a) In CPU; and (b) in GPU. Basically, CPU calculates a single incidence angle from a 3D point to a pinhole ((0,0) pinhole first, and (0,1) is the next one). On the contrary, GPU calculates every incidence angles at the same time thanks to GPU's parallelism.

and GPU.

We also exploited GPU technique into our approach in order to solve our drawbacks and reduce the processing time. In fact, we faced a critical defect of heavy-repetitive computation scheme at the integral image generation procedure. In our conventional approach, we calculate an incidence angle from a 3D point to a single pinhole (for example, the left-top positioned yellow pinhole $(0,0)$ in Figure 3.4(a)), and then move to the rest of the consecutive pinholes sequentially. This repetitive and time-intensive task takes a great deal of time because of a single thread from CPU manages these whole procedures. Thus, we moved to GPU's parallel computing methodology in order to boost the computation speed. We assigned each pinhole's index to GPU's threads and execute the incidence angle calculation all together in a parallel way. This approach is more effective and improves the performance dramatically. Figures 3.4(a) and (b) show the distinct approaches to calculate the incidence angles in CPU and GPU.

3.3 Display of 3D scenes using an integral-imaging monitor

In the last years, we have exploited normal 2D display devices, such as a liquid crystal display (LCD) monitor, a tablet PC, or even a mobile phone, in order to show an integral image and observe the displayed 3D scene. We could say that the important core parameters that has to be fixed to properly display the integral image are the following: the distribution of lenslets within the lens array (for instance, a quadrangle or honeycomb arrangement), the focal length of each lenslet, the number of pixels per millimeter of the display panel, the location of the lenslet centers respect to the middle of EIs, and the gap between the display screen and lens.

As we mentioned before, the display scheme in InI is a reversed sequence of the pick up procedure. We need to put an array of microlenses in front of the display panel with some specific gap, in order to restore the 3D information back from the composed integral image. If the gap is not equal to the focal distance of the lenslets, the displayed scene shows blurred information, or provides some color distortion problems. The size of EIs is also an important component at the display scheme. Each display device has its own specifications about how many pixels they have per millimeter (ppm). Most devices' ppm has decimal numbers. EIs computation when using this decimal number is more challenging than the use of a natural number. Furthermore, if the size of the displayed EIs does not fit with the lens pitch correctly, the adjacent pixels from neighboring EIs will invade each other. As a result, this size differentiation stops the visual perception of the 3D effect. For that reason, we firstly exploit natural number to count the number of EIs in the integral image generation scheme, and then we rescale the composed image to equalize the size between the display panel's ppm and EIs secondly.

In our experiments, we mainly used a Samsung SM-T700 tablet as our integral-imaging monitor. This device has a big number of ppm (14.1338px/mm), and most of all, this tablet fitted with our utilized MLA which has a square formation, equidistant position between lenses, and proper lens pitch (Model 630 from Fresnel Technologies, focal length f = 3.3 mm, pitch p = 1.0 mm). We mounted this MLA in front of our proposed display device and checked the displayed 3D scene.

To confirm our proposed experimental result, we composed an experimental set up, as shown in Figure 3.5. Originally, our main targets are binocular observers, who can see the 3D nature of displayed scene with their naked eyes. Unfortunately, the provided full-parallax effect cannot be directly demonstrated through a manuscript or even in a monocular video. To demonstrate this 3D effect, we replaced the binocular observer with a monocular digital camera, as a recording device. After that, a collection of pictures is obtained by displacing the camera in horizontal and vertical direction using a motorized linear stage in front of the proposed integral-imaging monitor. These captured pictures confirm that our proposed 3D monitor provides great parallaxes and dynamic viewing angles.

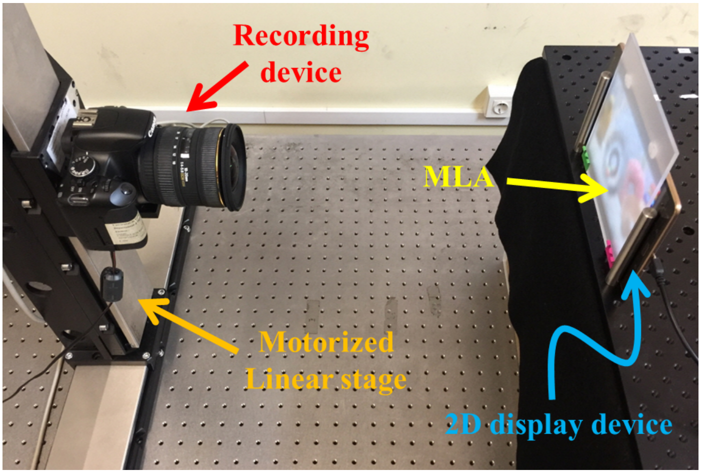


Figure 3.5: Overview of experimental system.

Chapter 4

Dense point cloud computation methods

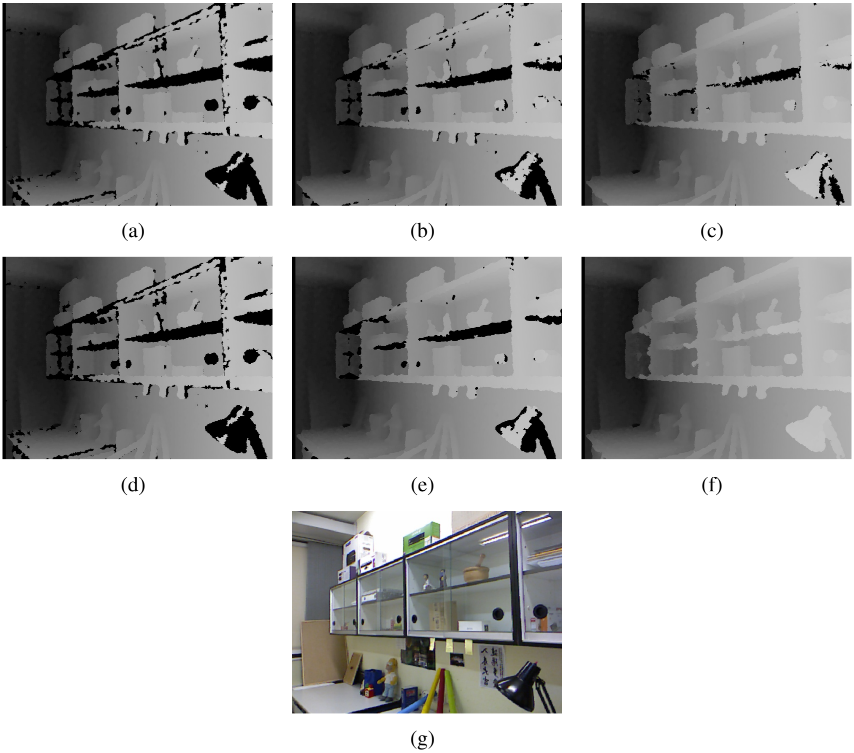
In this chapter, we will narrate our methods on how to improve the quality of 3D data, compose a distortion-free depth map from a captured plenoptic image, as well as recover the lost areas of the point cloud. As aforementioned, during last years, we mainly exploited the computerized 3D information at the moment of composing an integral image. Naturally, a dense point cloud not only assists to make a good quality of integral image, but also helps to provide an immersive 3D scene to the observers. On the other hand, a composed depth map contains several defects that are coming from various reasons. Among others, we faced several drawbacks that are coming from the utilized 3D camera's own limitation or its external problems, or loss of information from the occluded and/or concealed areas along the single camera's position. We will present our approaches and methodologies on how to solve such defects, and present our practical experimental results in following sub-sections. In the first section, we will provide our approach on how to restore the depth-hole areas (even the noisy areas) of the depth map image captured by the IR depth-sensing camera. In the following section, we will narrate a methodology on how to compose a dense depth map from a single-shot captured plenoptic image. We will finish this chapter by introducing the 3D data registration method for composing a dense point cloud, and then, providing our practical experimental results with some imaging experiments.

4.1 Depth-hole filtered point cloud

There are various types of 3D cameras in the market, and most of them are already utilized in many different research areas. Among them, the IR depth-sensing technique is one of the broadly used ones. However, the IR depth-sensing cameras have several drawbacks. In fact, these cameras acquire the depth images with some noises or depth-holes because of their own limitations and/or external factors. As we mentioned previously, the IR camera detects the projected IR light information (or emitted IR pulse in the case of the ToF camera system) and measures the depth distance. However, some materials will not reflect the IR light, or diverge the projected IR light to other directions due to the target object's non-planar shape. Besides, some transparent objects will not reflect but completely transmit the IR light. In some case, there are some interferences occurred that are coming from the different sourced ambient light, so that the IR depth-sensing camera cannot measure the depth distance properly. Several research groups tried to solve such drawbacks using their novel approaches [78, 79, 80, 81].

Among them, we mainly followed Camplani and Salgado's depth-hole filtering method [81] due to its good restoration performance and robustness. The strategy is the following. Firstly, register a base image for the algorithm to initialize the positions of target depth-hole areas and prepare for the restoration process. Secondly, capture other image to search the reliable depth information close to the depth-hole areas using the temporal-consistency map. Thirdly, collect the reasonable pixels that the registered image does not have, and vice versa. And lastly, fill in the certified pixels to the registered image, and update the filtering parameters. The algorithm iteratively computes the same sequence till the end of its iteration number. In fact, each frame of captured depth map image contains different information, and some depth-hole areas appear and disappear continually, so that the iterative approach is an adequate solution to restore the depth-hole areas. On the other hand, this depth-hole filtering structure utilizes two source images which are the depth map image and its correspond RGB image. It means that this algorithm considers the visual information in the depth map denoising procedure. In fact, the RGB image verifies the consistent adjacent pixels nearby the depth-hole areas correctly, and the use of visual information is a more effective way to alleviate the errors at the object boundaries of the depth map.

Besides, the joint-bilateral filter (JBF), which is an extended algorithm from the bilateral filter (BF), is applied in this filtering structure. The BF is a broadly used edge-preserved and noise-reduced smoothing filter. The weight of this filter is determined by the similarity of adjacent pixels, and non-similar neighboring pixels are not considered in the filtering procedure. Thus, the blurring effect near



(a) (b) (c)

(d) (e) (f)

(g)

Figure 4.1: Comparison result between the conventional depth-hole filtering algorithm and our proposed one: (a-c) Conventional algorithm's results; (d-f) proposed one's results; and (g) corresponding RGB image, respectively. Note that (a), (d) are in the initial frame, (b), (e) are 5 frames proceeded results, and (c), (f) are the results after 219 frames iteration. Through the panels, we can see the clear differences between them.

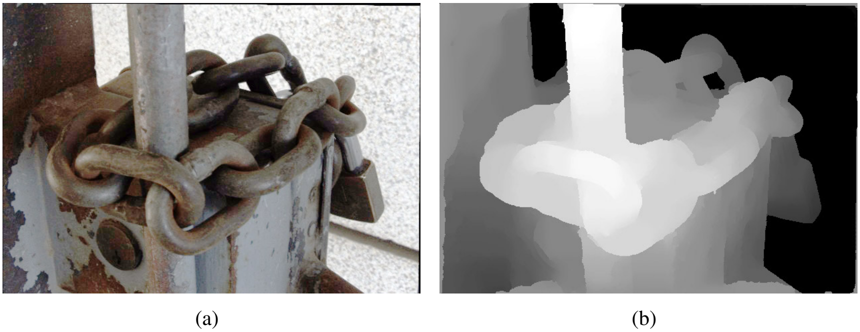
the boundaries is not much increased. The JBF has an extra weight from the BF that is selected by another guidance image, thus, this algorithm is also known as the cross-bilateral filter [82, 83]. So in sum, this depth-hole filtering algorithm restores the depth-hole pixels and homogeneous regions efficiently because of considering both, the RGB and depth images, and preserves the boundaries of objects thanks to the characteristic of JBF.

By the way, the conventional depth-hole filtering algorithm restores only very tiny areas per frame, so that it cannot cover the big sized depth-holes nor scattered small sized group of depthless pixels properly. But in fact, these small portions of depthless pixels are also parts of the potential recovering areas, and thus, the algorithm demands more further iterations to restore those regions. To solve such constraints, we additionally applied the median filter after the restoration phase of the certified pixels. The median filter considers each pixel in the image in turn and chooses the medium value between its adjacent pixels [80, 84]. Thanks to such characteristic of the median filter, it is efficient to expand the reliable depth values into their neighboring pixels, and clean up the noisy and meaningless pixels in object's boundary/edge regions. As a result, small and big depth-hole regions are recovered more efficiently (even faster) than the original algorithm. Figure 4.1 shows the comparison result between the conventional algorithm and our proposed one. In the conventional algorithm's result, some big depth-hole areas are not filled in properly, and even other small regions of depthless pixels still exist. On the contrary, the result of our proposed method shows that the depthless pixels are restored properly.

4.2 Dense point cloud computation from plenoptic camera

Plenoptic cameras have unique light gathering and post-capture processing capabilities. Thanks to such special merits, this camera is really spotlighted in recent years. Among others, one of the most widespread and interesting research tasks using plenoptic camera is dense depth map estimation from a captured plenoptic image. In fact, many research groups already announced their novel approaches and strategies [30, 31, 32]. But, we mainly followed Jeon's depth estimation method [32] due to the comprehensible depth estimation strategy, and the accessibility of modifiable data given by the authors.

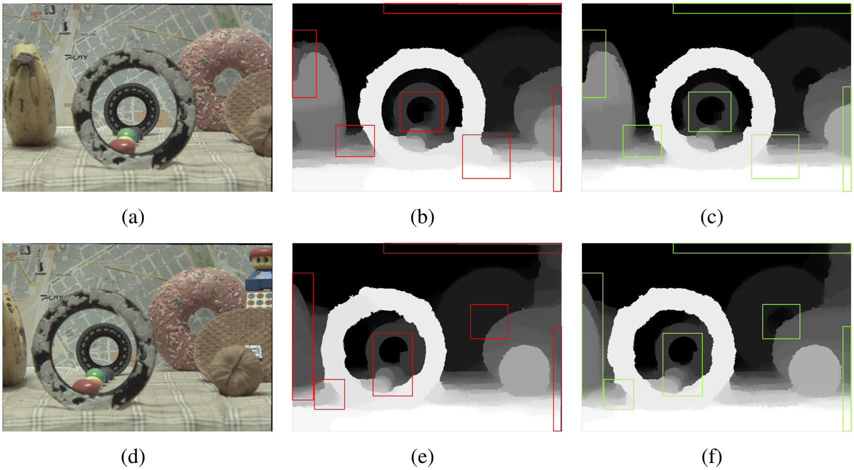
Jeon et al. compute the depth map using the stereo matching algorithm between sub-aperture images. Remarkably, the disparity range between adjacent sub-aperture images is very narrow (for instance, Lytro camera's disparity range is ± 1 pixel [85]), so that the sub-pixel accuracy calculation is strongly demanded. Due to such narrow disparity range, the shifting process is performed at the frequency domain, and the phase shift theorem is utilized. This scheme is an effective solution to displace an image position with an accurate sub-pixel distance. The shifted sub-aperture images become a foundation of the optimal disparities computation. After that, the stereo matching costs are computed between the central view image and other ones using a cost-volume-based stereo matching algorithm [86]. The computed per-pixel cost-volume is then refined using the weighted median filter [87] which is an edge-preserving filter and alleviates the coarsely scattered unreliable matches. Due to the very small viewpoint changes of the sub-aperture images, the feature correspondences between the central view image and other view images are used as an additional (or optional) constraint. Anyhow, with the refined cost-volume, the multi-label optimization process using the graph cuts algorithm [88] is propagated and the depth map is corrected at the texture regions, which are identified as being below satisfactory. Lastly, the fitting local quadratic function [89] is iteratively refining the computed local depth map to estimate a new non-discrete depth map. This procedure helps to solve the depth discontinuities effectively. Figure 4.2 shows our experimental result using the adopted depth map estimation algorithm. In our experiment, we firstly converted the captured plenoptic image to sub-aperture images array, and then computed the depth map with optimized parameters which are determined by several trials and errors.


(a)

(b)

Figure 4.2: Composed depth map image from a single-shot captured plenoptic image: (a) Center view image from the sub-aperture images array; and (b) composed depth map using Jeon's depth map estimation algorithm. We captured the image by using Lytro Illum plenoptic camera.

On the other hand, there were some unexpected image distortions that appeared in our captured plenoptic image. In fact, Jeon's approach considers an aspect of the MLA distortion problem, and it provides proper solutions in order to solve such drawback. However, the error that we found came from our camera's inherent performance (not related with the MLA distortions) so that we needed to solve it via another extra solution. We mainly adopted Dansereau's structure and utilized a given toolbox which is able to decode, calibrate, and rectify the lenselet-based plenoptic cameras through their specific procedures [66, 67]. After the plenoptic camera calibration, we performed the rectification procedure in each sub-aperture image, and finally, we could get the distortion-free depth map image. Figure 4.3 shows the comparison of depth map images provided by the conventional depth estimation method and our proposed one. The calibrated and rectified sub-aperture images compose a better quality of depth map images than the original ones. Especially, in the border areas of objects and the boundary of depth map images are displayed the clear differences. To the best of our knowledge, displaying a captured plenoptic image via the commercial plenoptic camera to the integral-imaging monitor was not addressed so far, and even not commonly handled such proposal before we performed.



(a) (b) (c)

(d) (e) (f)

Figure 4.3: Comparison result between the conventional depth estimation method and proposed one. Note that two image rows are the stereo-plenoptic image pair: (a-c) Left scene; and (d-f) right scene). On the other hand, each columns is the following: (a),(d) Center view image; (b),(e) conventional depth estimation method result; and (c),(f) after performing the plenoptic camera calibration with image rectification procedure, respectively. In fact, there are some unexpected image distortions at the border area of (a), (b), (d), and (e). On the contrary, (c) and (f) do not have any image distortion effect, and even they have bettered quality of depth map images. The red and green coloured rectangle areas indicate the clear differences between them.

4.3 3D data registration to restore the depth-holes using stereo camera systems

The image registration technique is a well-known task that overlays two or even more images of the same scene, which were captured from different perspectives by various imaging sensors and/or cameras. Thanks to the advancement of depth-sensing technology, the 3D data registration technique also got great attention, and various registration techniques were proposed by many research groups. According to the database of Institute of Scientific Information, more than 1000 papers were published on the topic of image and data registration during the last two decades [90]. Among others, the iterative closest point (ICP) algorithm is one of the broadly used techniques in order to fuse the 2D/3D data pairs [70, 71, 72]. The ICP algorithm aims to find the closest point on a geometric entity to a given point, and calculates the movement between data sets using an iterative refinement procedure. The output of ICP algorithm is the rigid (or rigid body) transformation matrix, which includes the translation and rotation information. However, non-rigid shapes are not allowed in the basic ICP algorithm. Due to such constraint, the non-rigid registration method is also spotlighted and actively researched till these days [91, 92, 93].

We applied the basic ICP algorithm into our research in order to get rid of the constraints of monocular vision system, and aimed to fill in the occluded and/or concealed areas of the composed point cloud. In fact, a single 2D/3D camera cannot avoid losing the overlapped areas or hidden surface information along the line of sight, and it is an inescapable limitation. On the contrary, multiple views enlarge the FOV and recover the occluded information by complementing each other. In our experiment, we set two differentiated stereo camera compositions and perform the registration procedure using composed point clouds: one is the stereo-hybrid 3D camera system, and the other is the stereo-plenoptic camera system.

In the stereo-hybrid 3D camera system, two depth sensing cameras were used, namely, one working with a structured IR pattern (Kv1) and another ToF 3D camera (Kv2). Both depth-sensing cameras have totally different characteristics, such as depth acquisition method, imaging sensor size (RGB and IR cameras of two devices, thus, totally 4 distinct imaging sensor sizes), captured image size, FOV, color tone of the acquired RGB images, etc. Thus, such different factors must be homogenized in order to make a well-unified point cloud set, and prevent a visual irregularity at the displayed 3D scene. Here, we adopted several significant algorithms to refine the heterogeneous point cloud pair. Firstly, the different image size problem between the two depth-sensing cameras is solved via the image

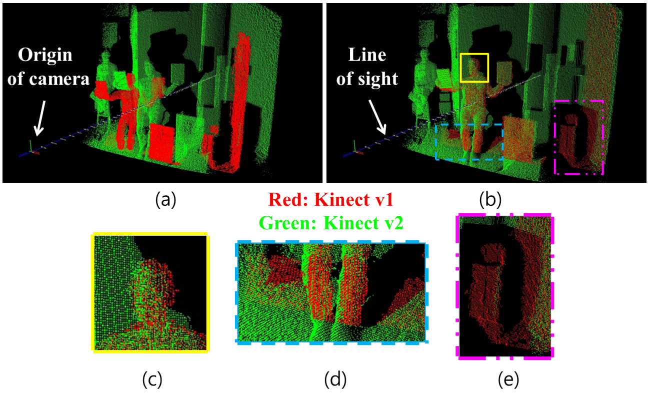


Figure 4.4: Simulated point cloud in the virtual 3D space: (a) and (b) Whole scene reconstruction before and after the registration process; and (c–e) some magnified specific parts of the registration result. The red coloured points are the composed point cloud from Kv1, and the green coloured ones are from Kv2. We can see that the occluded areas are restored properly thanks to the novel strategy for the uniformization process and the 3D registration technique. See text for further details.

scale correction method in hybrid stereoscopic camera system [94]. This method solves the different images scale information by considering various factors, such as, imaging sensor size, image size, focal length, FOV, and ppm. Next, the color tone dissimilarity between two RGB images is corrected by using the color transfer method [95]. Interestingly, this color correction method considers the color characteristic of both RGB images, and borrows one image's color characteristics from another. After correction of the major dissimilarities, two point clouds are fused via the basic ICP algorithm. Figure 4.4 shows the registration result between hybrid point clouds. In Figure 4.4(a), the red point cloud captured by Kv1 and the green one from Kv2 have each different scale and depth of volume, and even they are arranged irregularly. On the contrary, Figure 4.4(b) shows a well-arranged point cloud set thanks to the proposed unification strategy and the registration algorithm.

On the other hand, in the stereo-plenoptic camera system, we exploited Lytro Illum camera as a pick up camera, and the camera slider in order to capture a scene in two different positions. The camera slider not only assisted to move the camera's position easily, but also maintains the stability of capturing environment. Besides, it is difficult to arrange the 2 plenoptic cameras together with a small baseline, so

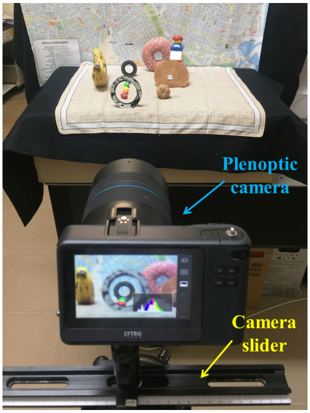
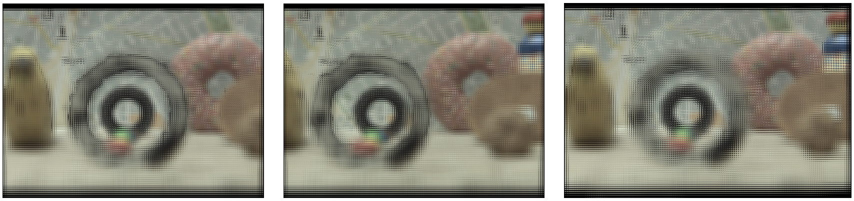


Figure 4.5: Overview of the proposed experimental environment for the stereo-plenoptic camera system. In this experiment, we exploit Lytro Illum plenoptic camera, and also the use of camera slider to acquire a different perspective view, generating an equivalent baseline that is rather small.

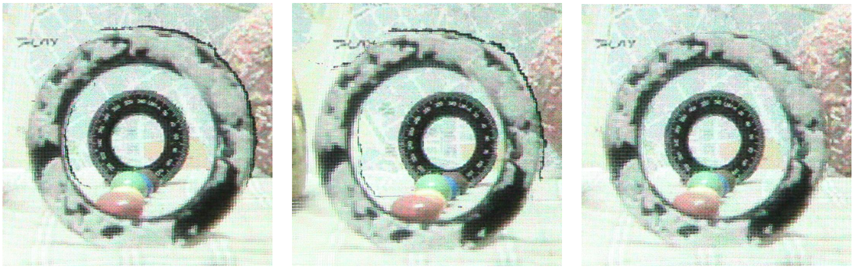
that the use of camera slider is a suitable alternative in such given situation. In this experiment, we do not need to perform any further uniformization process between a stereo image pair due to the advantage of using an identical camera. Firstly, we captured the first scene and moved to another position via the camera slider in order to acquire a different perspective scene, and computed a depth map image pair via our proposed approach, as in the previous section (Figures 4.3(c) and (f) are the composed results). The composed depth map image pair became two sets of point clouds, then, they were simulated in the virtual 3D space together. Finally, rigid transformation matrix was computed via the basic ICP algorithm and completed the registration process between the two point clouds.

Thanks to the stereo configuration, some occluded and lost areas of the point cloud are restored properly in both stereo camera systems. To confirm our proposal, we set the experimental environment as shown in Figure 4.5, and Figure 4.6. The figures present three different integral images composed by the different point cloud conformations: left and right scene, and the registered point cloud set, respectively. In our experiment, we registered the right scene's point cloud into the space of left one. The main reason is that the right scene not only contains the occluded information of the left scene, but also new objects appear at the scene. Finally, we displayed the composed integral images into our proposed integral-imaging monitor, and performed a comparison between them. To assess our comparison result, we recorded the displayed integral images at the same position, and excerpted specific common regions of the displayed 3D scene. In Figures 4.7(a) and (b), there are some black coloured areas at the displayed scene. These black areas are coming from the empty space, that is depth-holes or shadowing areas, so that these pixels have associated with meaningless depth information. On the contrary, thanks to the complementation between the two-point clouds, the black coloured regions are restored and covered precisely, as shown in Figure 4.7(c).



(a) (b) (c)

Figure 4.6: Composed integral images: (a) and (b) From left and right scene; and (c) fused one from (b) to (a), respectively. As we can see through the sub-figures, (c) has more abundant information and larger FOV than (a) and (b), thanks to the stereo camera system and registration algorithm.



(a) (b) (c)

Figure 4.7: Comparison between the displayed integral images: (a) and (b) Left and right scenes; and (c) from our proposed method's result. We clipped-out a common area of the recorded scenes in order to ease the comparison. In (a) and (b), there are some black coloured areas behind of objects. On the contrary, (c) does not have any black pixels.

Chapter 5

Summary of Papers

In this chapter, we will summarize the contributions of our 6 main papers, which are the basis of this thesis. Each paper provides various and distinct approaches on how to compose and refine the dense 3D data acquired by the different types of cameras. The papers also narrate our techniques and solutions on how to provide an immersive 3D experience to the observers via our proposed integral-imaging monitor. We will describe each paper's main proposals and achievements concisely, but clearly. The papers appear in the chronological order of publication, and their contents are organized as follows. In Paper I, our first approach on how to compose an integral image from the captured 3D data is presented. In Paper II, a methodology on how to restore the depth-hole areas in the captured depth map image by using an efficient hole-filtering algorithm is explained. In Paper III, a new type of depth-sensing camera is exploited into our experiment, and then the comparison of performances between the conventionally used 3D camera and newly adopted one is preceded via several components and factors. In Paper IV, the usage of stereo 3D camera configuration which is composed by two different types of depth-sensing cameras at the occluded/concealed areas restoration procedure is explained. The strategy on how to uniformize the different elements of heterogeneous cameras is also narrated. In Paper V, a new approach on how to compose an integral image with point cloud in order to extend our conventional method is presented. Lastly, in Paper VI, the usage of stereo-plenoptic camera system, which is a novel approach in InI, at the dense 3D data composition phase, is explained. Also, a methodology on how to boost the integral image computation time using GPU acceleration technique is narrated.

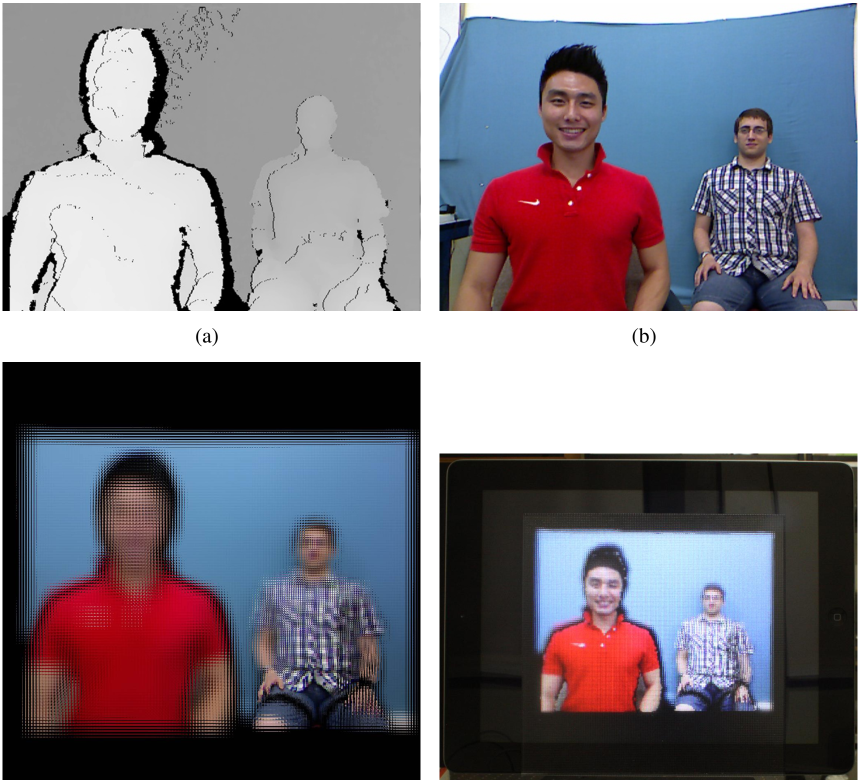
5.1 Paper I

A conventional display system is only able to provide 2D information despite of the original object is 3D shape. In fact, there are a lot of depth-sensing techniques to acquire dense 3D information, whereas their final displayed scenes are converted to a plane 2D image. Here, we mainly focused on such paradoxical situation, in which the digitized 3D information is not displayed in their original volumetric structure through a conventional display system.

There are many alternatives to get a kind of 3D experience by using the glasses-type stereoscopic 3D system (anaglyph, active or passive 3D glasses), or a wearable device (HMD). However, these systems need that the observers put on these equipments to their heads obligatorily to get the 3D perception. Besides, such systems only provide a single perspective view so that the displayed scene seems to be too artificial and to be far apart from the real experience. Furthermore, the vergence-accommodation conflict inherent in these architectures leads to visual fatigue in the mid run.

To avoid such drawbacks, we exploited autostereoscopic 3D technology, especially InI technique, which provides an immersive 3D perception, continuous viewing points, full-parallax, and also presents the full-colored scenes to multiple observers properly. It is noteworthy that the observers can see the 3D scene with their naked eyes directly, without using any additional wearable device or equipment. Additionally, we mainly adopted the depth-sensing camera into our experiment in order to compose a dense depth map of the 3D scene in real-time. Then, we simulated the computerized 3D volumetric information into our own experimental system in order to restore the 3D scene as it is.

Our main contribution in this paper is to compose a PS-free integral image by using the captured 3D information and the InI technique. To restore and display the acquired 3D scene as its original volumetric distribution, our proposed integral-imaging monitor is utilized. Firstly, a dense 3D scene is acquired via the IR depth-sensing camera, and a set of point cloud is composed (Kv1 is mainly utilized in this paper). The point cloud is composed by the combination between a depth map and its corresponding RGB image. Secondly, the computerized point cloud is simulated into the virtual 3D space, and the VPA, whose parameters are configurable, is placed in a proper position at will. In fact, we put the VPA's position inside of (or close to) the simulated point cloud. Thirdly, the projection scheme is followed from each 3D coordinate point to all virtual pinholes, one after another. The methodology on how to compose the EIs by using the VPA is already illustrated in Chapter 3. Fourthly, the composed EI is rotated by 180°about its center position in order to avoid PS problem, as following Okano's approach. And lastly,



(a)

(b)

(c)

(d)

Figure 5.1: The first experimental result from our proposed method: (a) Single-shot captured depth map image from Kv1; (b) corresponding RGB image; (c) composed integral image using the VPA; and (d) displayed integral image via the proposed integral-imaging monitor. Note that the position of the VPA (also the reference plane in this configuration) is positioned in the right human model's middle thigh (see Figures (c) and (d)). Thus, the left human model is floating outside from the monitor.

the composed integral image is resized in order to fit with the physical MLA's pitch correctly, and then the produced image is displayed via our proposed 3D monitor. However, the provided full-parallax effect cannot be directly demonstrated in a manuscript, because of the original target of proposed 3D display system is for binocular observers. To prove the effectiveness of our experimental implementation, we captured the displayed scene through a monocular digital camera moving in horizontal and vertical direction using a motorized linear stage. Note that such demonstration method is mainly adopted in our further papers due to the clarity of proof and ease of use.

To the best of our knowledge, the procedure of displaying a captured 3D scene from a depth-sensing camera into an integral-imaging monitor was not addressed so far at that time, what confirms the novelty of our research. Our approach only needs a single depth map and its corresponding RGB image, so that the requirements to produce an integral image and display the full-parallax 3D scene are very simple. Besides, the way of placing the VPA close to the point cloud has the great merit of being able to select a reference plane of the displayed scene in a much simpler and more intuitive way (note that the position of VPA and the reference plane's one are equal in this configuration). But over all, the proposed methodology becomes the foundation of this thesis, and the beginning of our whole proposed researches.

5.2 Paper II

Kv1 is well-known for its capacity of capturing both, the RGB and the depth map images simultaneously. However, the IR and RGB cameras of Kv1 are physically distanced each other so that they must be mapped from one's view to another. To solve such drawback, we exploited initially the Kinect's software development kit provided by Microsoft. Unfortunately, this method only supports a mapping procedure from the IR camera to RGB camera, and does not permit to map in the opposite sequence (from the RGB to IR camera). Besides, there is some noise appearing after the mapping procedure. That drawbacks must be solved.

On the other hand, Kv1 uses a structured IR light pattern emitter, and the IR camera captures and measures the depth distances through the acquired pattern information. However, this type of depth-sensing camera cannot avoid capturing a defective depth map image because of its own limitations and/or external factors. There are some groups of black coloured pixels in the acquired depth map image that correspond to depth-holes or depthless pixels, which do not contain any depth values. Such meaningless pixels are appearing due to some of the following problems: occlusion, limited capacity of the depth distance acquisition range, relative surface angle, or even surface materials. As a result, the depth-holes deteriorate the quality of composed integral image and eventually relieves the 3D perception.

In this paper, we aimed to fuse the views of RGB and IR cameras of Kv1

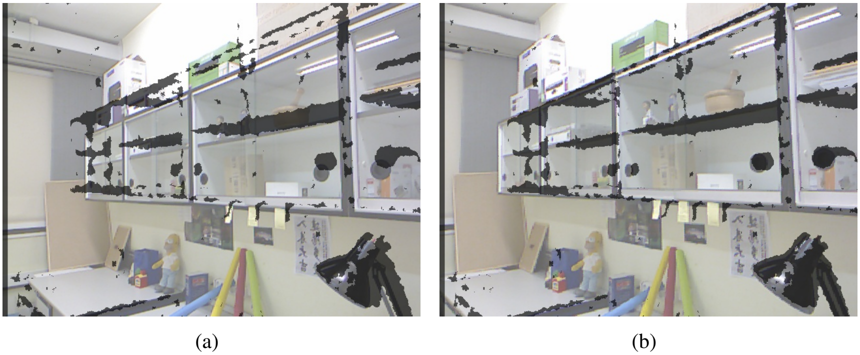


Figure 5.2: The comparison of the raw mapping result between the RGB and depth map image: (a) Raw images mapping result; and (b) mapping result after the calibration procedure. In this experiment, we were able to map the RGB image to the depth map image. Note that such mapping sequence is not provided by the given software development kit from Microsoft. On the contrary, our proposed approach is able to map between two cameras at will.

without using the given software development kit. We also performed the restoration process of depth-holes in order to improve the quality of displayed 3D scene. Firstly, the camera calibration is performed in order to calculate and estimate the inherent parameters of the RGB and IR cameras (the intrinsic and extrinsic parameters). Then, the coordinate systems of both cameras are mapped by using the calibrated parameters, and finally, the RGB and depth map images are merged properly. Note that we applied the progressive threshold image accumulation method, and a chessboard pattern is utilized in the camera calibration phase, in order to find the correspond feature points between two images easily. Secondly, Camplani and Salgado's depth-hole filtering method is adopted in order to fill in and recover the depth-holes. This depth-hole filtering algorithm restores the depth-hole pixels efficiently after considering both, the RGB and depth images information together. We already explained this depth-hole filtering algorithm in Chapter 4 with details. Thirdly, a dense point cloud is composed through the combination of the refined depth map image and its corresponding RGB image, and eventually, the integral image is computed and displayed via the integral-imaging monitor.

The camera calibration process between both, the RGB and IR cameras from Kv1, allows to expand the limited usage of the original software from the manufacturer. The calibrated parameters help to map between two cameras freely, at will.

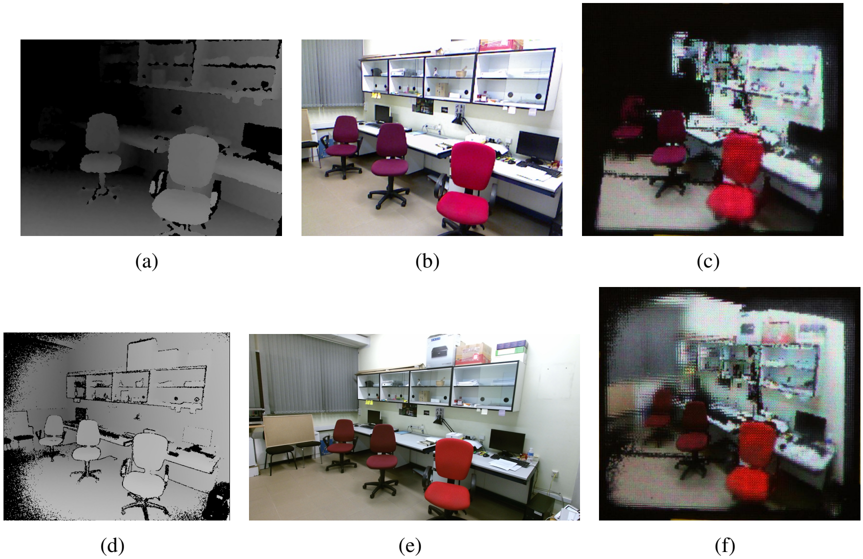
Our proposed depth-hole filtering algorithm upgrades the performance of adopted hole-filtering algorithm in a more efficient way (see Figure 4.1 in Chapter 4). Finally, the refined depth map assists to compose an improved quality of the integral image, and also helps to display a more immersive 3D scene to the observers.

5.3 Paper III

There are various types of commercialized depth-sensing cameras in the market. Among them, Kv1 and Kv2 from Microsoft have been really spotlighted and widely applied in many different research areas during the recent years. In fact, both 3D cameras have totally different features, such as, depth acquisition method, captured image size, FOVs of the RGB and IR cameras, etc. On the other hand, although the commercial specifications are announced from the manufacturer, several detailed information are veiled and not provided to the consumers, like as, the coupled areas of the scene between the RGB and IR cameras, FOV of the RGB camera, and the density of the captured depth map, etc. For that reason, we aimed to analyze such veiled properties and the inherent capacities of the exploited depth-sensing cameras, and then confirm and verify the known specifications through our experimental results. We also wanted to expand our conventional research approach, so that we applied totally different types of 3D cameras into our experiment, and compared the results between the conventionally used IR depth-sensing camera and newly adopted one.

In this paper, we set several experimental setups to compare the performances and characteristics of Kv1 and Kv2. Firstly, coupled areas of the scene captured by the RGB and IR cameras of Kv1 and Kv2 are verified and analyzed: where they are and how much areas are sharing together. Here, we use the chessboard pattern as a common target, and then the captured images from the RGB and IR cameras are merged into a single image through the computed correlation parameters. Secondly, the FOVs of the RGB and IR cameras from Kv1 and Kv2 are calculated via the empirical parameters, and then the FOVs information from the commercial specification is compared with our derived results. Thirdly, the RGB and depth map images are captured by Kv1 and Kv2 separately, and then two sets of point clouds are composed and simulated into the virtual 3D space. After that, the density of depth information and detailed matters are compared and analyzed. Lastly, two different integral images are displayed via our proposed 3D monitor, and the final comparison results are presented.

The confirmed and verified issues through our experiments are the following. Firstly, the coupled areas of the scene between the RGB and IR cameras from Kv1

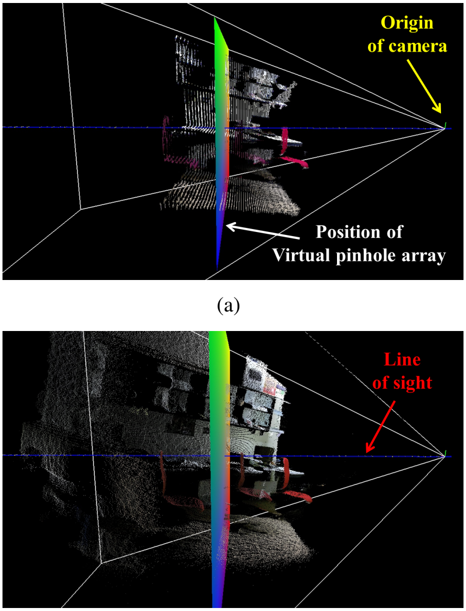


(a) (b) (c)

(d) (e) (f)

Figure 5.3: Comparison result between Kv1 and Kv2: (a) and (b) Depth map and RGB images captured by Kv1; (d) and (e) from Kv2, respectively. (c) and (f) are excerpted images of the displayed 3D scenes from our proposed 3D display system. These two distinct results confirm the obvious differences between Kv1 and Kv2.

and Kv2 are checked and analyzed (see Figure 2.6 in Chapter 2). Secondly, we confirmed that the announced FOV information from the commercial specification given by the manufacturer is only for the IR camera. Besides, the reliability of official parameters are checked and proved again by us. Thirdly, the density of acquired depth information from Kv1 and Kv2 have different formations and characteristics (see Figure 5.4). Kv1 has specific layered structures in each certain depth distances, so that it is an unavoidable defect that appears some depthless layers in the point cloud continuously. On the contrary, the Kv2 provides denser depth information without any regularized figuration, whereas there are some depthless pixels appeared at times in each corners of the captured depth map image (Figure 5.3(d)). Lastly, the displayed 3D scene (Figures 5.3(c) and (f)) certified that the Kv2 has much bettered lateral of depth, and having a long range of axial distances than the Kv1.



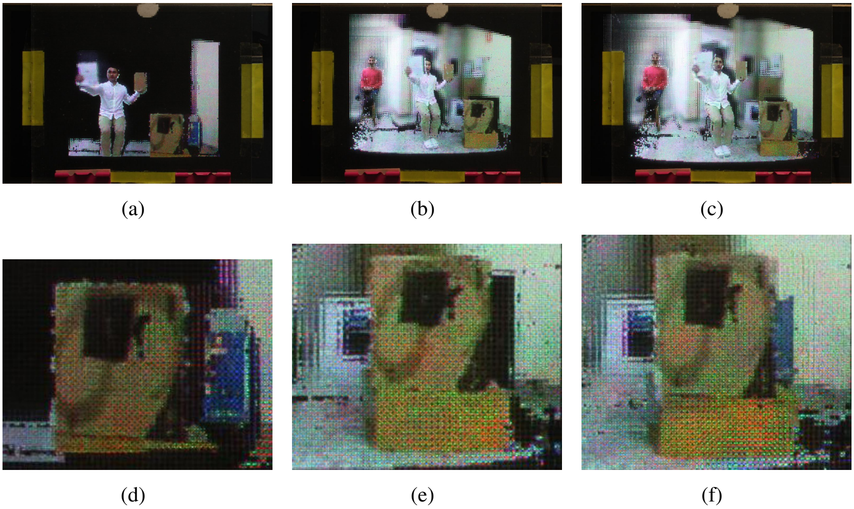
(a)

(b)

Figure 5.4: Display of the point clouds into the virtual 3D space: (a) From Kv1; and (b) from Kv2. In both cases, the position of VPA is located just behind of the second chair. Interestingly, there are some specific gaps and empty layers in the simulated 3D points in (a), but not in (b). The depth volume of (b) is also deeper than (a) because of the different capacity of depth acquisition between Kv1 and Kv2.

5.4 Paper IV

A single 2D/3D camera cannot avoid losing the information of overlapped areas or hidden surfaces along the line of sight, and it is an inescapable defect of mono perspective view. On the contrary, a multiple camera composition enlarges the FOV and recovers the occluded information by complementing each other. Due to such benefits, we exploited the multiple cameras into our experiment. The main purpose of our experiment is to compose a stereo 3D camera set up, and to fuse two sets of point clouds in order to fill in the vacant volumetric areas. In fact, the stereo configuration is the most fundamental and basic approach for the multiple camera system. Here, we utilized heterogeneous 3D cameras into our experiment. The main reason is that there are many different 3D cameras in the market, and we could combine and apply the various different brands of 3D cameras in our further additional experiments. However, two different depth-sensing cameras must be homogenized due to their heterogeneous properties, so that further unification processes are strongly demanded.



(a) (b) (c)

(d) (e) (f)

Figure 5.5: Comparison result between the displayed integral images: (a) and (b) Captured by Kv1 and Kv2; and (c) composed result from our proposed approach. (a) has lack of visual information due to the limited depth acquisition capacity of Kv1, whereas it contains different perspective information from (b). On the contrary, (b) has more abundant information than (a) thanks to the bettered depth acquisition performance of Kv2. However, several objects are concealed and occluded (for instance, the blue box in (a) and (d)). After the registration process between (a) and (b), several depth-hole areas are filled in and restored properly thanks to the stereo 3D camera configuration (the registration process is performed from (a) to (b) sequence). To compare between the results, we excerpted the common area of displayed scenes (d-f).

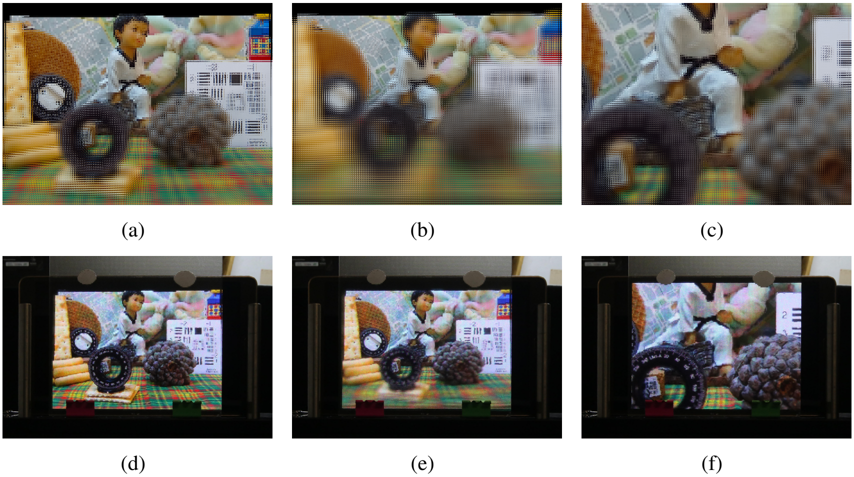
In this paper, we exploited Kv1 and Kv2. Both depth-sensing cameras have totally different characteristics, thus they are suitable equipments according to the concept of our research. Firstly, the different image size problem between two depth-sensing cameras is solved via the image scale correction method. Secondly, the color tone dissimilarity between two captured RGB images is corrected by using the color transfer method. Thirdly, the homogenized two sets of point clouds are then fused via the basic ICP algorithm. Note that such unification strategy is already explained in Chapter 4 with details. Lastly, the fused point cloud is simulated into the virtual 3D space, and then a bettered integral image is composed and displayed via the proposed 3D monitor.

To the best of our knowledge, this was the first time to utilize the stereo-hybrid 3D camera system to capture the light field, what confirms the novelty of our research. Our proposed experimental system allows recovering the occluded/concealed volumetric areas efficiently thanks to the stereo 3D camera configuration and the registration algorithm. The proposed unification strategy also helps to homogenize the heterogeneous point clouds pair efficiently. We demonstrated and illustrated the comparison results between hybrid cameras by using various imaging experiments and details (see Figure 5.5). The last but important thing is that further researches are available by exploiting the combination of the different brands of heterogeneous depth-sensing cameras in a future.

5.5 Paper V

Thanks to the advance of science and technology, some proposals of capturing and transmitting images in real-time were accomplished with a great deal during the past two decades. After all, several companies announced their plenoptic cameras (or light field cameras), which were influenced by Lippmann's IP theory. The main merit of these cameras is capturing both, the spatial and angular information of light rays at the same time, proceeding from the natural scene in 3D. The manufacturers also provide their handy solutions to extract good qualities of the RGB and depth map images through the given softwares.

In this Paper, we exploited the plenoptic camera in order to expand our research area by using a new type of camera. Besides, we applied the conventional method in order to compose an integral image by using the point cloud, as in Paper 1. However, as we already mentioned in Chapter 3, this method has a drawback. If a 3D point from a simulated point cloud is positioned really close to the VPA (in front of or behind of the pinholes array), it cannot be projected to other pinholes because of the limited incidence angles. Even, the neighboring 3D points also cannot be penetrated to other adjacent pinholes, so that such situation causes to form some apparent vacant areas in the composed integral image. To solve this defect, we adopted the concept of synthetic aperture method into our experiment.



(a) (b) (c)

(d) (e) (f)

Figure 5.6: Composed integral images with different factors: (a) Composed integral image through the empirically determined optimum factors; (b) pitch factor changed from (a); and (c) FOV modified from (a). (d), (e), and (f) are the displayed 3D scenes through the proposed integral-imaging monitor. The composed integral images have a same reference plane (positioned in the surface of a wall). (b) and (e) present the blurred scenes because of 5 times larger pitch than (a). (c) and (f) provide the magnified scenes of (a), in which the FOV is 2 times narrower than (a).

Our proposal in this paper can be divided into 5 steps. Firstly, a point cloud is composed by the combination between the acquired RGB and depth map images, and simulated into the virtual 3D space. Here, we directly extracted the RGB and depth map images by using the given software from the manufacturer (we exploited Lytro Illum plenoptic camera from Lytro company). Secondly, the VPA is displaced far away from the point cloud, as in Figure 3.3(b) in Chapter 3. Note that the number of virtual pinholes is set equal to the number of pixels which is displayed behind of the each lenslet of the integral-imaging monitor. Besides, the position of VPA is empirically decided in order to consider the trade-off between the resolution of EIs and the position that the black coloured pixels appear in the captured images. For instance, if the VPA locates too close to the point cloud, the composed scene cannot cover the entire information, and even some vacant spaces (shadowing areas of the captured objects) will appear at the scene. In contrast, if the VPA moves further from the simulated point cloud, the resolution of the composed EIs are decreased, so that the process of finding a proper position of the VPA must be underwent following trials and errors. But in the end, the final position of the VPA, the cropping factor, and the shifting factor will be define after the decision of the reference plane's location and the region of interested views of the scene. Thus, the position of VPA and reference plane's one are distinct and separate components in this configuration, contrary to Paper 1. Thirdly, each sub-aperture image is clipped out by considering the cropping factor and shifting factor. In fact, the cropping factor selects the size of cropping area at the captured EIs, and the shifting factor decides the moving distance from center-view's cropping region to its neighboring views' one. Fourthly, the cropped sub-aperture images are resized in order to match with the size of the spatial resolution of integral-imaging monitor. And fifthly, the resized sub-aperture images are converted to the integral image, as following Figure 2.3 in Chapter 2.

Our proposed approach helps to avoid the black pixels appearing in the integral image, and also assists to compose a large DOF at the displayed scene. Thanks to the modifiable factors, the FOV is freely selectable and zoom in a specific region of interest area of the scene is available, at will (see Figures 5.6(c) and (f)). On the other hand, it is difficult to decide the optimum parameters for the best result. If we consider a larger gap between pinholes, the DOF is reduced and the computed integral image provides a blurred image at the displyed scene (see Figures 5.6(b) and (e)). On the contrary, if the gap is smaller, the depth perception is reduced and the parallaxes are mutually decreased; as a result, the reconstructed 3D scene has flat depth volume and eventually loses the 3D sensation. That's why it must be underwent many trials and errors in order to find not only the adequate parameters but also the reference plane and the VPA's proper position.

In sum, as we mentioned in Chapter 3, the proposed method in this paper is appropriate to be exploited when the displayed 3D scene not only needs a great DOF and uniformly-focused scene, but also demands certainly selected reference plane. On the contrary, our conventional method, as in Paper 1, is suitable to be applied when the displayed scene needs great parallaxes, abundant depth sensation, and the necessity of tuning the reference plane's position frequently.

5.6 Paper VI

As we mentioned previously, the plenoptic camera has great merits of transcribing various information into a single-shot captured image. Thanks to such novelty, this camera has been spotlighted by many photographers and consumers, and even many scientific researchers also had a great interest of its potential possibilities.

Among all, this camera got big attention due to the possibility of extracting a dense depth map. In fact, the manufacturers provide their solutions to extract a good quality depth map image from a captured plenoptic image, and also support various useful functions to modify the scene through the given software, but unfortunately, they did not open their technical data. Due to such veiled techniques, many research groups tried to compute and estimate the depth map by using their novel approaches and solutions.

Meanwhile, hardware acceleration techniques are broadly applied nowadays throughout the whole society at large. As we mentioned in Chapter 3, GPU accelerated computing, which is a representative hardware acceleration technology, is mainly aimed to process some vast amounts of data quickly or performs heavy computations more efficiently. Due to such benefits and great performances of GPU, we also studied this trendy and powerful technique, and eventually adopted into our research.

In this paper, we illustrated our proposal to compose a dense point cloud from a pair of single-shot images captured by the stereo-plenoptic camera configuration. We exploited GPU acceleration technique in order to boost the heavy-repetitive computation scheme in the integral image generation procedure. Firstly, a pair of plenoptic images is captured by the proposed stereo-plenoptic camera system. As we explained in Chapter 4, we utilized the camera slider in order to capture the scenes in each different position in an easier way, and also to maintain the stability of the capturing environment. In fact, it is difficult to arrange the 2 plenoptic cameras together in a narrow baseline, thus, the use of camera slider is a suitable alternative in such given situation (see Figure 4.5 in Chapter 4). Moreover, thanks to the use of a single camera in the experiment, it was not necessary to perform any further unification procedures. Secondly, two dense depth map images were computed, and then two point clouds were composed. In the depth map image estimation phase, we mainly followed Jeon's method due to the comprehensible depth estimation strategy (Figures 4.2, 4.3(b), and (e) in Chapter 4 present the composed results). Meanwhile, some unexpected image distortions appeared in the computerized RGB and depth map images, so that we firstly performed the plenoptic camera calibration with the rectification procedure in both captured plenoptic images, and then secondly computed the depth map images (Figures 4.3(c) and (f) in Chapter 4 show the results). The pair of depth map images and RGB images were finally modified into two dense point clouds, and simulated into the virtual 3D space. Thirdly, these two point clouds were fused by using the ICP algorithm, as in Paper 4. Fourthly, the integral image was composed and the computation time was boosted through GPU acceleration technique. In fact, in our conventional approach, we calculated an incidence angle from a 3D point to a single pinhole, and then moved to the rest of consecutive pinholes sequentially. This repetitive and time-intensive task takes a great deal of time. Thus, we assigned each pinhole's index to GPU's threads and executed the incidence angle calculation all together in a parallel way. Fifthly, the composed integral image was displayed and the comparison result between the singular plenoptic cameras and proposed stereo camera configuration were presented (see Figures 4.6 and 4.7 in Chapter 4).

To the best of our knowledge, this was the first time to utilize the stereo-plenoptic camera system to compose a dense point cloud and display the captured scenes with full-parallax, what confirms the novelty of our research. The defects of tilt and distortion problems in the captured plenoptic image were solved via the adopted plenoptic camera calibration and rectification methods. The proposed stereo-plenoptic camera configuration and the adopted registration algorithm help to recover the occluded/concealed volumetric areas efficiently. The last but important thing is that we boosted the integral image computation time by exploiting GPU acceleration technique. Figure 5.7 presents the comparison result between CPU and GPU's computation time clearly.

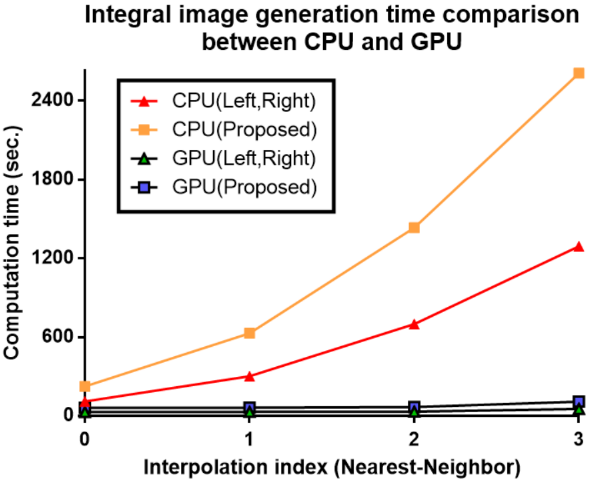


Figure 5.7: Comparison of integral image computation time between CPU and GPU. The triangles represent the left and right scene's computation results, and the rectangles indicate the fused point cloud's integral image computation results, respectively. As shown in this figure, GPU accelerated computation speed is much faster than CPU's computation result when performing both tasks.

Chapter 6

Conclusions

In this thesis, we devoted the best endeavors to produce an immersive sense of depth and perception of 3D, via the combination of computerized 3D information and InI technique. Our proposals in this thesis are simple and concise, but they have originality. To the best of our knowledge, our contributions were not addressed so far at that time, and even not commonly handled before we performed.

To begin with, we simulated a set of composed point clouds into the virtual 3D space, and put an array of virtual pinhole cameras nearby (inside or a little way off) the point cloud, at will. After that, the displaced pinholes array composed the synthetic information with as many as the numbers of virtual cameras and with tunable parameters. The composed images are then properly handled and edited, and eventually integrated into an integral image which contains the spatial and angular information at the same time. Finally, the produced image is displayed in our proposed 3D monitor, providing an immersive depth perception and full-parallax to multiple observers all together. Note that this methodology became the foundation of this thesis, and also the beginning of our whole proposed researches.

Meanwhile, we proposed several approaches on how to handle and refine the 3D data acquired by different types of cameras. In fact, we faced some issues in the computerized depth map, which are coming from the inherent problems of the utilized cameras, or even due to external factors. Moreover, a singular camera cannot avoid losing the information of overlapped areas, or the occurrence of occluded surfaces along the line of sight, so that it is an unavoidable defect of mono perspective view. To solve such limitations and/or inescapable problems, we adopted and applied various novel approaches and techniques. Note that some algorithms were improved and upgraded respect to the original performance because of our additional supplements. The introduction of stereo camera configuration enlarged the FOVs and contributed to recover the occluded information by complementing each other. Here, the composed pair of point clouds is integrated into a single point cloud properly through the adopted algorithm, and eventually, a bettered quality integral image is composed and displayed. The last but important thing is that we solved the critical defect of heavy-repetitive computation at the integral image generation phase via a hardware acceleration technique. Thanks to the parallel computing methodology, we improved the efficiency and also boosted the computation speed dramatically.

As a final remark, we would like to introduce and comment about our further possibilities of research. Actually, there are many different 3D cameras in the market currently, so that we could combine and apply the various different brands of depth-sensing cameras into our further experiments. Of course these adopted cameras must be homogenized due to their heterogeneous properties, so that a lot of diverse and different unification procedures should be also strongly demanded. The combination between microscopy and our proposals will be a great research task and also a part of a potential application area in the future work. On the other hand, applying various registration methods in the multiple 3D camera composition is another potential research complement. In fact, we adopted the basic registration algorithm into our experiment, but as a matter of fact, there are a lot of robust algorithms proposed by several research groups. Thus, our potential research complements will also be to improve the quality of displayed 3D scene by applying diverse registration methods.

Hopefully, we would mind that our achievements of the combination between computerized dense 3D information and InI technology could be a piece of milestone to anyone interested in and also needed. We also desire that our efforts would help the mass adoption of these technologies via various research areas in the coming years.

Bibliography

[1] E. Kheirandish, *"The Arabic Version of Euclid's Optics"*, 1st edition, Springer-Verlag New York Inc. (1999).

[2] Sir D. Brewster, *"The stereoscope"*, John Murray, London (1856).

[3] M. Lahanas, "Ancient Greece: Optics", `http://www.hellenicaworld.com/Greece/Science/en/Optics.html` (Consulted on September 16th, 2019).

[4] R. du Fresne, *"Trattato della Pittura di Lionardo da Vinci"*, Giacomo Langlois, Paris (1651), `https://archive.org/details/gri_33125008484301/page/n7` (Consulted on September 16th, 2019).

[5] Sir C. Wheatstone, *"Contributions to the physiology of vision.—Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision"*, Philos. Trans. R. Soc. Lond. 128, 371-94 (1838).

[6] W. Rollmann, *"Notiz zur Stereoskopie"*, Ann. Phys. 165, 350–1 (1853).

[7] W. Rollmann, *"Zwei neue stereoskopische Methoden"*, Ann. Phys. 166, 186–7 (1853).

[8] R. Zone, *"Stereoscopic Cinema and the Origins of 3-D Film, 1838-1952"*, 64-66, The University Press of Kentucky (2007).

[9] M. Brain, "How 3-D Glasses Work", `http://www.howstuffworks.com/3-d-glasses2.htm` (Consulted on September 16th, 2019).

[10] T. L. Turner and R. F. Hellbaum, *"LC shutter glasses provide 3-D display for simulated flight"*, J. Inf. Disp. 2, 22-24 (1986).

[11] L. Edwards, “Active Shutter 3D Technology for HDTV”, `http://phys.org/news173082582.html` (Consulted on September 16th, 2019).

[12] L. Edwards, “Active shutter 3D system”, `http://en.wikipedia.org/wiki/Active_shutter_3D_system` (Consulted on September 16th, 2019).

[13] C. Demers and M. Azzabi, “3D TVs: ACtive 3D vs Passive 3D”, `https://www.rtings.com/tv/learn/3d-tvs-active-3d-vs-passive-3d` (Consulted on September 16th, 2019).

[14] B. Andrén, K. Wang, and K. Brunnström, *“Characterizations of 3d tv: active vs passive”*, SID Symp. digest of technical papers, 43, 137-40 (2012).

[15] I. E. Sutherland, *“A head-mounted three dimensional display”*, Proc. ACM Joint Computer Conference (AFIPS ’68) (Fall, part I), 757-64 (1968).

[16] M. G. Tomilin, *“Head-mounted displays”*, J. Opt. Technol. 66, 528-33 (1999).

[17] G. Lippmann, *“Épreuves réversibles photographies intégrals”*, Comptes Rendus de l’Académie des Sciences 146, 446–51 (1908).

[18] G. Lippmann, *“Épreuves réversibles donnant la sensation du relief”*, J. Phys. Theor. Appl. 7, 821–25 (1908).

[19] H. Arimoto and B. Javidi, *“Integral three-dimensional imaging with computed reconstruction”*, Opt. Lett. 26, 157–9 (2001).

[20] S. Manolache, A. Aggoun, M. McCormick, N. Davies, and S. Y. Kung, *“Analytical model of a three-dimensional integral image recording system that uses circular- and hexagonal-based spherical surface microlenses”*, J. Opt. Soc. Am. 18, 1814–21 (2001).

[21] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, *“Lightfield microscopy”*, ACM Trans. Graph. 25, 924–34 (2006).

[22] E. H. Adelson and J. Y. A. Wang, *“Single lens stereo with plenoptic camera”*, IEEE Trans. Pattern Anal. Mach. Intell. 14, 99–106 (1992).

[23] T. Georgiev and A. Lumsdaine, *“The focused plenoptic camera and rendering”*, J. Electron. Imaging 19, 021106 (2010).

[24] F. Okano, H. Hoshino, J. Arai, and I. Yayuma, *"Real time pickup method for a three-dimensional image based on integral photography"*, Appl. Opt. 36, 1598–603 (1997).

[25] H. Navarro, R. Martínez-Cuenca, G. Saavedra, M. Martínez-Corral, and B. Javidi, *"3D integral imaging display by smart pseudoscopic-to-orthoscopic conversion (SPOC)"*, Opt. Express 18, 25573-83 (2010).

[26] M. Martínez-Corral, A. Dorado, H. Navarro, G. Saavedra, and B. Javidi, *"Three-dimensional display by smart pseudoscopic-to-orthoscopic conversion with tunable focus"*, Appl. Opt. 53, 19-25 (2014).

[27] J-H. Jung, J. Kim, and B. Lee, *"Solution of pseudoscopic problem in integral imaging for real-time processing"*, Opt. Lett. 38, 76–8 (2013).

[28] M. Levoy and P. Hanrahan, *"Light Field Rendering"*, Proc. 23rd Ann. Conf. Comput. Graph. Interact. Technol. 31-42 (1996).

[29] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, *"The Lumigraph"*, Proc. 23rd Ann. Conf. Comput. Graph. Interact. Technol. 43-54 (1996).

[30] N. Sabater, M. Seifi, V. Drazic, G. Sandri, and P. Pérez, *"Accurate disparity estimation for plenoptic images"*, Proc. Eur. Conf. on Comput. Vis. 548–60 (2014).

[31] C. Huang, *"Robust pseudo random fields for light-field stereo matching"*, IEEE Conf. Comput. Vis. Pattern Recogn. 11-19 (2017).

[32] H. Jeon, J. Park, G. Choe , J. Park, Y. Bok , Y. Tai, and I. Kweon, *"Accurate depth map estimation from a lenslet light field camera"*, IEEE Conf. Comput. Vis. Pattern Recogn. 1547–55 (2015).

[33] G. Scrofani, J. Sola-Pikabea, A. Llavador, E. Sanchez-Ortiga, J. C. Barreiro, G. Saavedra, J. Garcia-Sucerquia, and M. Martínez-Corral, *"FIMic: design for ultimate 3D-integral microscopy of in-vivo biological samples"*, Biomed. Opt. Express 9, 335–46 (2018).

[34] M. Cho and B. Javidi, *"Three-dimensional tracking of occluded objects using integral imaging"*, Opt. Lett. 33, 2737–39 (2008).

[35] H. Bae, J. Kim, and J-P. Heo, *"Content-Aware Focal Plane Selection and Proposals for Object Tracking on Plenoptic Image Sequences"*, Sensors 19, 48 (2019).

[36] A. Dorado, M. Martínez-Corral, G. Saavedra, and S. Hong, *"Computation and display of 3D movie from a single integral photography"*, J. Disp. Technol. 12, 695-700 (2016).

[37] Bumblebee2, FLIR Systems, `https://www.flir.com/products/bumblebee2-firewire/` (Visited on September 16th, 2019).

[38] ZED, Stereolabs, `https://www.stereolabs.com/` (Visited on September 16th, 2019).

[39] T. Kanade and M. Okutomi, *"A stereo matching algorithm with an adaptive window: theory and experiment"*, IEEE Trans. Pattern Anal. Mach. Intell. 16, 920-32 (1994).

[40] Middlebury College, "Stereo Datasets - Middlebury Computer Vision", `http://vision.middlebury.edu/stereo/data/` (Consulted on September 16th, 2019).

[41] FLIR Systems, "How is depth determinded from a disparity image?", `https://www.flir.com/support-center/iis/machine-vision/knowledge-base/how-is-depth-determined-from-a-disparity-image/` (Consulted on September 16th, 2019).

[42] Z. Zhang, *"Flexible camera calibration by viewing a plane from unknown orientations"*, Proc. IEEE 7th Int. Conf. on Comput. Vis. 666–673 (1999).

[43] Z. Zhang, *"A flexible new technique for camera calibration"*, IEEE Trans. Pattern Anal. Mach. Intell. 22, 1330–1334 (2000).

[44] OpenCV: camera calibration, OpenCV, `http://docs.opencv.org/doc/tutorials/calib3d/camera_calibration/camera_calibration.html` (Consulted on September 16th, 2019).

[45] Kinect v1, Microsoft, `https://support.xbox.com/en-US/xbox-on-windows/accessories/kinect-for-windows-info` (Visited on September 16th, 2019).

[46] Kinect v2, Microsoft, `https://support.xbox.com/en-US/xbox-on-windows/accessories/kinect-for-windows-v2-info` (Visited on September 16th, 2019).

[47] Canesta sensor, Canesta, `https://en.wikipedia.org/wiki/Canesta` (Visited on September 16th, 2019).

[48] Calmine 1.08 and 1.09, PrimeSense, `https://en.wikipedia.org/wiki/PrimeSense` (Visited on September 16th, 2019).

[49] Xtion Pro, Asus, `https://www.asus.com/3D-Sensor/Xtion_PRO/` (Visited on September 16th, 2019).

[50] Realsense, Intel, `https://software.intel.com/en-us/realsense` (Visited on September 16th, 2019).

[51] "Microsoft Kinect 'fastest-selling device on record'", BBC, `https://www.bbc.com/news/business-12697975` (Consulted on September 16th, 2019).

[52] D. Pagliari and L. Pinto, *"Calibration of kinect for Xbox one and comparison between the two generations of microsoft sensors"*, Sensors 15, 27569–89 (2015).

[53] R. Smeenk, "Kinect v1 and Kinect v2 fields of view compared", `http://smeenk.com/kinect-field-of-view-comparison` (Consulted on September 16th, 2019).

[54] A. Shpunt and Z. Zalvesky, *"Depth-Varying Light Fields for Three Dimensional Sensing"*, US. Patent 8, 50, 461, 8, May (2008).

[55] K. Khoshelham, *"Accuracy Analysis of Kinect Depth Data"*, Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 38, 133-138 (2011).

[56] J. Sell and P. O'Connor, *"The XboxOne System in a Chip and Kinect Sensor"*, IEEE Micro. 34, 44–53 (2014).

[57] G. Bradski and A. Kaehler, *"Learning OpenCV: Computer Vision with the OpenCV Library"*, 381-383, O' Reilly Media Press, California (2008).

[58] R. Sukthankar, R. G. Stokton, and M. D. Mullin, *"Smarter Presentations: Exploiting Homography in Camera-Projector Systems"*, Proc. Int. Conf. Comput. Vis. 247-253 (2001).

[59] S. Hong, Y. Tan, H. Yeo and B-G. Lee, *"1-inch UniTouch System using Kinect"*, Int. Conf. on Signal Proc. Img. Proc. Pattern Recogn. 351-355 (2013).

[60] R. Ng, *"Digital Light Field Photography"*, PhD dissertation, Stanford University, Stanford, CA (2006).

[61] PiCam: Pelican Imaging Camera, Pelican Imaging, `http://lightfield-forum.com/light-field-camera-prototypes/pelican-imaging-array-camera-light-field-module-for-smartphones/` (Consulted on September 16th, 2019).

[62] K. Venkataraman, D. Lelescu, J. Duparre, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar, *"Picam:An ultra-thin high performance monolithic camera array"*, ACM Trans. Graph. 32, 2504-2507 (2013).

[63] Lytro camera, Lytro, `https://en.wikipedia.org/wiki/Lytro` (Visited on September 16th, 2019).

[64] Ratrix camera, Ratrix, `https://www.raytrix.de` (Visited on September 16th, 2019).

[65] K. Ohashi, K. Takahashi, M. P. Tehrani and T. Fujii, *"Super-Resolution Image Synthesis Using the Physical Pixel Arrangement of a Light Field Camera"*, IEEE Int. Conf. Img. Proc. 2964-2968 (2015).

[66] D. G. Dansereau, O. Pizarro, and S. B. Williams, *"Decoding, calibration and rectification for lenselet-based plenoptic cameras"*, IEEE Conf. Comput. Vis. Pattern Recogn. 1027–1034 (2013).

[67] D. G. Dansereau, "Lightfield toolbox for matlab", `https://dgd.vision/Tools/LFToolbox` (Consulted on September 16th, 2019).

[68] M. Levoy, M. Gross, and H. Pfister, *"Chapter 2: The early history of point-based graphics" in Point-Based Graphics*, 9-16, Eds. Burlington, MA, USA: Elsevier (2007).

[69] M. Levoy and T. Whitted, *"The Use of Points as a Display Primitive"*, TR 85-022, University of North Carolina at Chapel Hill (1985).

[70] P. J. Besl and N. D. Mckay, *"A method for registration of 3-D shapes"*, IEEE Trans. Pattern Anal. Mach. Intell. 14, 239–256 (1992).

[71] Z. Zhang, *"Iterative point matching for registration of free-form curves and surfaces"*, Int. J. Comput. Vis. 13, 119–152 (1994).

[72] S. Rusinkiewicz and M. Levoy, *"Efficient variants of the ICP algorithm"*, Int. Conf. on 3-D Digit. Imaging and Model. 145–152 (2001).

[73] M. Martínez-Corral, B. Javidi, R. Martínez-Cuenca, and G. Saavedra, *"Formation of real, orthoscopic integral images by smart pixel mapping"*, Opt. Express 13, 9175–9180 (2005).

[74] J. Grepstad, "Pinhole Photography-History, Images, Cameras, Formulas", `https://jongrepstad.com/pinhole-photography/pinhole-photography-history-images-cameras-formulas/` (Consulted on September 16th, 2019).

[75] J-S. Jang and B. Javidi, *"Three-dimensional synthetic aperture integral imaging"*, Opt. Lett. 27, 1144–1146 (2002).

[76] J. D. Owens, D. Luebke, N. Govindaraju, M. Harris, J. Krüger, A. E. Lefohn, and T. J. Purcell, *"A survey of general-purpose computation on graphics hardware"*, Comput. Graph. Forum 26, 80-113 (2007).

[77] D. Tarditi, S. Puri, and J. Oglesby, *"Accelerator: Using Data Parallelism to Program GPUs for General-Purpose Uses"*, SIGOPS Oper. Syst. Rev. 40, 325-335 (2006).

[78] A. Butler, S. Izadi, O. Hilliges, D. Molyneaux, S. Hodges, and D. Kim, *"Shake'n'sense: Reducing interference for overlapping structured light depth cameras"*, Proc. of SIGCHI. 1933-1936 (2012).

[79] A. Criminisi, P. Perez, and K. Toyama, *"Region filling and object removal by exemplar-based image inpainting"*, Image Proc. IEEE Trans. 13, 1200–1212 (2004).

[80] S. Matyunin, D. Vatolin, Y. Berdnikov, and M. Smirnov, *"Temporal filtering for depth maps generated by kinect depth camera"*, 3DTV-CON art. no. 5877202, 1–4 (2011).

[81] M. Camplani and L. Salgado, *"Efficient spatio-temporal hole filling strategy for Kinect depth maps"*, Proc. SPIE Int. Conf. 3-D Image Process. Appl. 1–10 (2012).

[82] C. Tomasi and R. Manduchi, *"Bilateral filtering for gray and color images"*, Int. Conf. Comput. Vis. 846, 839–846 (1998).

[83] G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. Cohen, and K. Toyama, *"Digital photography with flash and no-flash image pairs"*, Proc. of SIGGRAPH, 664–672 (2004).

[84] T. Huang, G. Yang and G. Tang, *"A fast two-dimensional median filtering algorithm"*, IEEE Trans. Acoust., Speech, and Signal Process. 27, 13-18 (1979).

[85] Z. Yu, X. Guo, H. Ling, A. Lumsdaine, and J. Yu, *"Line assisted light field triangulation and stereo matching"*, Int. Conf. Comput. Vis. 2792-2799 (2013).

[86] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, *"Fast cost-volume filtering for visual correspondence and beyond"*, IEEE Conf. Comput. Vis. Pattern Recogn. 3017-3024 (2011).

[87] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu, *"Constant time weighted median filtering for stereo matching and beyond"*, Int. Conf. Comput. Vis. 49-56 (2013).

[88] V. Kolmogorov and R. Zabih, *"Multi-camera scene reconstruction via graph cuts"*, Proc. Eur. Conf. Comput. Vis. 82-96 (2002).

[89] Q. Yang, R. Yang, J. Davis, and D. Nistér, *"Spatial-depth super resolution for range images"*, IEEE Conf. Comput. Vis. Pattern Recogn. 18-23 (2007).

[90] B. Zitová and J. Flusser, *"Image registration methods: a survey"*, Image Vis. Comput. 21, 977–1000 (2003).

[91] H. Chui and A. Rangarajan, *"A new point matching algorithm for non-rigid registration"*, Comput. Vis. Image Underst. 89, 114-41 (2003).

[92] D. Hähnel, S. Thrun, and W. Burgard, *"An extension of the ICP algorithm for modeling nonrigid objects with mobile robots"*, Int. Joint. Conf. Artif. Intell. 915-20 (2003).

[93] J. Ma, J. Zhao, J. Jiang, and H. Zhou, *"Non-rigid point set registration with robust transformation estimation under manifold regularization"*, Proc. Conf. Artif. Intell. 4218–24 (2017).

[94] H. Shin, S. Kim, and K. Sohn, *“Hybrid stereoscopic camera system”*, J. Broadcast Eng. 16, 602–13 (2011).

[95] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, *“Color transfer between images”*, IEEE Comput. Graph. Appl. 21, 34–41 (2001).

Papers I-VI

Paper I

Towards 3D Television Through Fusion of Kinect and Integral-Imaging Concepts

Seokmin Hong, Donghak Shin, Byung-Gook Lee, Adrián Dorado,
Genaro Saavedra, and Manuel Martínez-Corral

in *Journal of Display Technology* 11, 894-9 (2015).

Towards 3D Television Through Fusion of Kinect and Integral-Imaging Concepts

Seokmin Hong, Donghak Shin, Byung-Gook Lee, Adrián Dorado, Genaro Saavedra, and Manuel Martínez-Corral

***Abstract*—We report a new procedure for the capture and processing of light proceeding from 3D scenes of some cubic meters in size. Specifically we demonstrate that with the information provided by a kinect device it is possible to generate an array of microimages ready for their projection onto an integral-imaging monitor. We illustrate our proposal with some imaging experiment in which the final result are 3D images displayed with full parallax.**

***Index Terms*—Integral imaging, kinect, 3D monitors.**

I. Introduction

CONVENTIONAL photography is fully adapted for recording in a 2D sensor the images of the 3D world. Although the images produced by photography are essentially 2D, they carry many cues that account for the 3D nature of the recorded scenes. This is the case, among others, of the perspective rules, which make closer objects to appear bigger than further ones. This effect is due to the well-known fact that the size of the image in the photographic sensor is determined by the angular size of objects. Other cues are shadows, occlusions, or defocus. In case of video recording, the relative speed of moving objects (or static objects when the camera is moving) is also a significant depth cue. For most of applications, the capture and display of 2D images provides enough information and/or satisfaction to users and minimizes the amount of data to be stored, transmitted and displayed. This is the reason for the still massive use of 2D photography and video.

However, the need for capturing and displaying the 3D information of 3D scenes is increasing very fast in the 21st Century. Its potential application in, for example, microscopy [1], [2], medical imaging [3]–[6], optical inspection in production chains [7], security monitoring [8], or virtual simulators for civil or military applications [9], etc., makes the capture and display of 3D images a hot topic in the research end/or engineering for the next decade.

If we discard, at this moment, holography, which still needs coherent illumination, or stereoscopy, which does not provide real 3D experiences, we can affirm that technology based on the Integral Photography principle is in the right way of producing acceptable 3D experience. Integral Photography was proposed, in 1908, by Gabriel Lippmann [10]. His proposal intended to face the problem of conventional photographic cameras which, when working with 3D scenes, do not have the ability of recording the angular information carried by the rays of light passing through their objective [10]. Instead, the irradiance received by any pixel is proportional to the sum of radiances of all the rays, regardless of their incidence angle. To overcome this lack, Lippmann proposed to insert a microlens array (MLA) in front of the photographic film. This permits to register an array of microimages, which store a radiance map with the spatial and angular information of all the rays proceeding from the 3D scene.

The radiance map has been named in different ways, such as integral photography [10], integral imaging [12], lightfield map [13] or even plenoptic map [14], [15]. From this map it is possible, for example, to tackle the challenge of displaying 3D images with a flat monitor [16]–[19].

As for the methods for the capture of the integral imaging, some application-dependent proposals have been made along the past few years. When the 3D scene is small and close to the camera, the plenoptic architecture, in which the MLA is inserted at the focal plane of the camera lens, seems to be the best adapted [20],[21]. Also interesting is the use of a small array of tiny digital cameras [22], which can be inserted in a cellular phone. However, due to low parallax, its utility is restricted to close objects.

When the 3D scenes are much bigger, of the order of some cubic meters, a different capture rig is necessary. In this case, the most useful proposal have been based on the use of large camera arrays, arranged either in 1D or in 2D grid [23], [24]. Note that in this case, the proposed techniques need an extremely accurate synchronization between the cameras, and make use of a huge amount of data, which are unnecessary for display purposes.

Manuscript received August 28, 2014; revised November 09, 2014; accepted November 11, 2014. Date of publication November 13, 2014; date of current version October 05, 2015. This work was supported in part by the Plan Nacional I+D+I (grant DPI2012-32994), Ministerio de Economía y Competitividad (Spain), by Generalitat Valenciana (Spain) under project PROMETEOII/2014/072, and by the IT R&D program of MKE/KEIT [10041682].

S.-M. Hong, D.-H. Shin and B.-G. Lee are with the Institute of Ambient Intelligence, University of Dongseo, 617-716 Busan, Korea (e-mail: tommyhsm@gmail.com; shindh2@gmail.com; lbg@gdsu.dongseo.ac.kr).

A. Dorado, M. Martínez-Corral and G. Saavedra are with the Department of Optics, University of Valencia, E-46100 Burjassot, Spain (e-mail: adrian.dorado@uv.es; genaro.saavedra@uv.es; manuel.martinez@uv.es).

This paper has supplementary material available at http://www.ieeexplore.ieee.org, provided by the author. This pack contains two videos (Media 1 and Media 2). Media 1.mov shows a 3D scene, composed by two office chairs, as displayed by an integral-imaging monitor. Media 2.mov shows a 3D scene, with a human model, as displayed by an integral-imaging monitor. The file is 15.4 MB in size.

Color versions of one or more of the figures are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JDT.2014.2370734

Our aim here is to propose the fusion between two concepts that are very different, but which are very successful in the area of 3D imaging and sensing. We refer to integral imaging and to kinect technology. This kind of fusion was proposed previously, but with different aim [25]. Kinect technology permits the registration, in real time, of accurate depth maps of big, opaque, diffusing 3D scenes. This is obtained with low resolution, which, however, matches perfectly with the requirements of resolution of integral-imaging monitors. Then, we propose first to capture the sampled depth map of a 3D scene with the Kinect. Second, simulate with our software, the capture of the sampled depth map with an array of digital cameras whose position, pitch and resolution are in good accordance with the characteristics of the integral-imaging monitor. And third, to project this information onto the monitor, so that the lenses of the MLA integrate the light emitted by the pixels, producing 3D scenes displayed with continuous perspective and full parallax.

II. Acquiring 3D Points Cloud With Kinect

Although, as stated above, there are different methods to record the information of a three-dimensional scene, in this contribution we used a Kinect device, which was initially launched, by Microsoft[1], as an add-on accessory for the Xbox game console on 2010. Its unique features have been determinant to find applications in human's full body tracking [26], motion detection and voice recognition. However, the distinctive hallmark of this device is its capability for recording simultaneously the RGB image and the depth information in real-time. This can be made because the Kinect has two different cameras, which can operate with the same resolution [27], a RGB camera and an infra-red (IR) one. The principle behind the Kinect technology is based on depth mapping obtained from projected structured IR patterns. The Kinect's IR emitter projects a fixed pattern onto the target and both the depth distance and the 3D reconstructed map are obtained from the reflected pattern recorded by the IR camera [28], [29]. The depth information provided by the Kinect is ranged between 800–6000 mm from the sensor plane. However, data should be acquired, generally, between 1000 and 3000 mm. This is due to the quality degradation of the depth data for larger distances as result of the noise and its own low resolution [30].

Our aim here is to achieve a point cloud that includes the information of 3D position and color intensity. Although the Kinect provides such information, a limitation comes out from the fact that the two cameras (RGB and IR) are physically separated from each other and, therefore, their fields of view are different. Consequently, both scenes will not match properly, see left picture in Fig. 1.

To overcome this drawback, we use the function named 'NuiImageGetColorPixelCoordinateFrameFromDepthPixelFrameAtResolution' which is provided by the software development kit (SDK) of the Kinect, from Microsoft.[2]. This function operates by matching depth information onto RGB image and it works in real time. Fig. 1, right picture, shows the final result after the proper matching between two cameras.

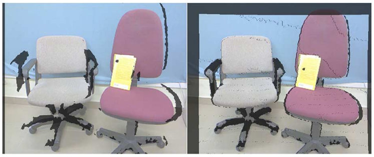


Fig. 1. (left) Raw mapping result and (right) reconstructed mapping result after our proposal method. See text for further details.

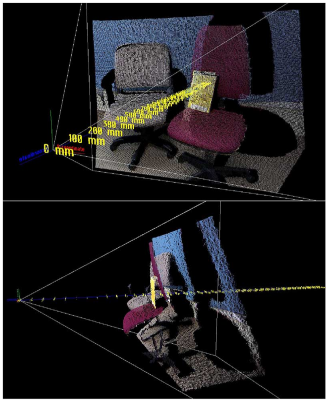


Fig. 2. Display of a 3D points-cloud. From both panels, it is clear that each point is given by its (x, y, z) position and its color intensity.

After matching the two images, we reassign the information into points located in a 3D virtual space using OpenGL environment. Now, each point is defined by six values: its (x, y, z) coordinates and its RBG color intensities. Based on their depth positions, the correct arrangement of the whole points is satisfied by using Standard Template Library (STL) and its associative container that is called 'multimap'.[3], [4]. In Fig. 2, the generated 3D point cloud is shown.

III. Depth Arrangement of the 3D Points-Cloud

The next step of our procedure is to prepare the algorithm for the calculation of the microimages for their projection onto the integral-imaging monitor. To this end we need first to express the spatial 3D coordinates of the points in a homogeneous way. Take into account that coordinates of the 3D point cloud produced after the previous section, are expressed in pixels (x and y coordinates) and in millimeters (z coordinate). To make the system homogeneous we performed a calibration experiment,

[1] Microsoft Kinect. [Online] Available: http://www.xbox.com/en-us/kinect/

[2] Kinect SDK. [Online] Available:http://msdn.microsoft.com/en-us/library/hh855347.aspx

[3] STL. [Online] Available: http://en.wikipedia.org/wiki/Standard_Template_Library

[4] STL-map. [Online] Available: http://msdn.microsoft.com/library/1fe2x6kt(v=vs.110).aspx

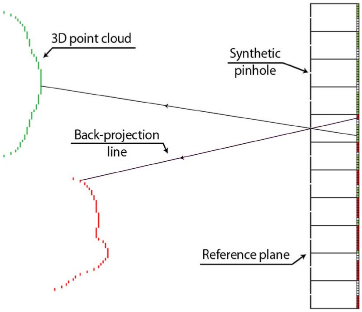


Fig. 3. Scheme of the algorithm for calculating the microimages. The number of pixels of the integral image and the number of (x,y) pixels of the points cloud are similar.

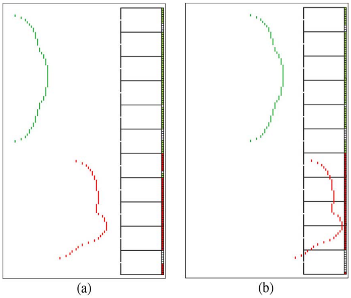


Fig. 4. Scheme of the back-projection algorithm. (a) The reference plane is close to the 3D point cloud; (b) The reference plane is within the 3D cloud. Pixel assignment strongly depends on the cloud position.

using a chessboard as the object, and concluded that in our case one pixel was equal to two millimeters in the object space.

For the second step, the characteristics of the InI monitor need to be known and expressed in pixel coordinates. Specifically, in our experiment we used an iPad equipped with retina display (264 pixels/inch), and a MLA consisting of 147 × 147 lenslets of focal length $f_L = 3.3$ mm and pitch $p = 1.0$ mm (Model 630 from Fresnel Technology). Then, for our algorithms, any microimage was composed by 11 pixels, the gap between the microlenses and the display was fixed to $g = 36.3$ px, and therefore the full size of the integral image would be at most 1617 × 1617.

IV. Micro-Images Generation

To generate the microimages we first resize laterally the 3D points cloud from 480×640 pixels to 1213×1617 pixels. Note that this change does not produce any distortion, since the ratio is still 4:3. In our computer calculation we simulate an experiment of capture of microimages. In this experiment we placed the points cloud at a certain distance from a simulated pinhole array. The distance is equal to the distance between the original scene and the kinect (see Fig. 3). Note that from now and hereafter, the plane where the synthetic pinhole array is placed will be named as the reference plane. Then we assigned the values of the pixels of the microimages by back-projection through the pinholes, as in [31].

Note that when these microimages are projected onto the monitor and displayed through the microlenses, the result will be a 3D image that is floating at a big distance from the monitor. This can reduce drastically the resolution of the displayed 3D scene, and provide a windowed aspect to it. What is more convenient is to prepare a set of microimages such that the displayed 3D image is in the neighborhood of the MLA, with some parts in front of it and some other parts behind. To obtain this directly with our algorithm, we can shift axially the 3D cloud towards the synthetic pinhole array, so that the reference plane is within the 3D scene (see Fig. 4). From the figure it is apparent that the microimages strongly depend on the reference plane position.

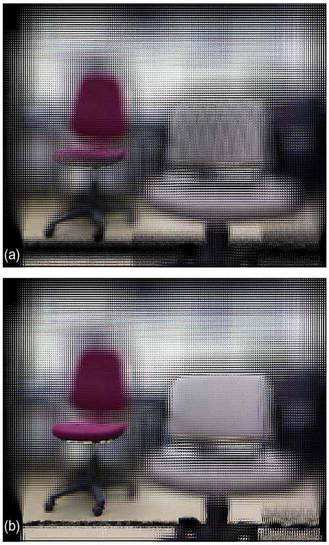


Fig. 5. Collection of microimages generated from the 3D points cloud captured with the kinect. (a) The full scene is in front of the pinhole array, like in the scheme shown in Fig. 3. (b) The 3D points cloud was displaced toward the pinhole array. In this case the front part of the seat of the red office chair is at the pinhole-array plane.

Finally, following Okano [32], we rotate any microimage by 180° about its center to avoid the pseudoscopic display, and resize the matrix to 1145 × 1527 to take into account the resolution of the retina display (10.39 px/mm).

V. Experimental Results

First we show, in Fig. 5, the microimages calculated for two different positions of the reference plane.

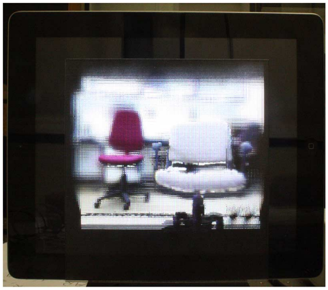
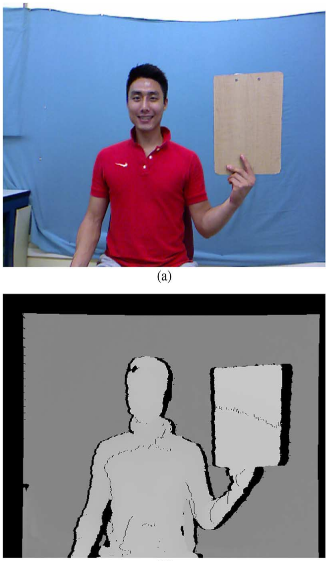


Fig. 6. Single-frame excerpt from video recording of the implemented InI monitor (Media 1). The video is composed of views of the monitor obtained from different horizontal and vertical positions.

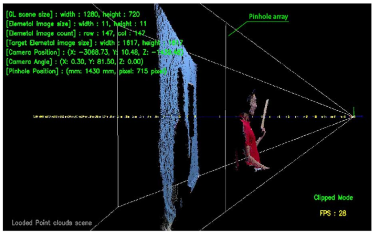


(a)

(b)

Fig. 7. Kinect output for a 3D scene with human model. (a) RGB picture. (b) Depth map from the IR picture.

Then, the microimages shown in Fig. 5(b) were displayed on iPad. The MLA was properly aligned so that pixels were close to the focal plane. Exact adjustment could not be made, due to the transparent plate that covers the retina display. This small misadjustment, about 0.5 mm, resulted in some braiding effect [33]. The InI monitor is shown in Fig. 6. It is apparent from the movie that the monitor projects a full parallax 3D image. This has been possible after a single shot capture thanks to the fusion between the kinect capture and integral-imaging processing and display.



```
[GL scene size] : width : 1280, height : 720
[Elemetal Image size] : width : 11, height : 11
[Elemetal image count] : row : 147, col : 147
[Target Elemetal image size] : width : 1617, height : 1617
[Camera Position] : (X: -3068.73, Y: 10.48, Z: -1430.48)
[Camera Angle] : (X: 0.30, Y: 81.50, Z: 0.00)
[Pinhole Position] : (mm: 1430 mm, pixel: 715 pixel)
```

Clipped Mode

FPS : 28

Loaded Point clouds scene

Fig. 8. 3D points cloud of the captured scene. In green text we show the parameters for the microimages calculation and the position of the reference (or pinhole-array) plane.

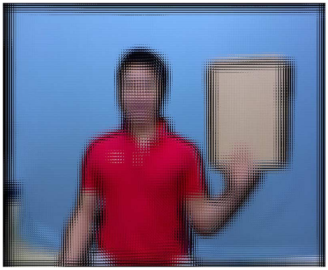


Fig. 9. Integral image ready for its projection onto the InI monitor. Note that the reference plane was set just behind the back of the human model.

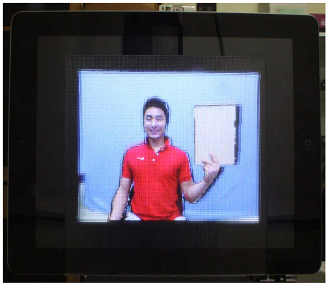


Fig. 10. Single-frame excerpt from video recording of the implemented InI monitor (Media 2).

To confirm the utility of our approach we did a second experiment and applied the procedure to a 3D scene with a human model. In Fig. 7 we show the RGB and the depth-map images obtained with the kinect. From this information we calculated the 3D points cloud shown in Fig. 8. In such figure we show also the parameters for the microimages calculation. Finally, in Fig. 9 we show the microimages obtained after application of our algorithm, which are ready for their projection onto the retina display of the iPad.

Again we aligned the MLA, just placed in contact with the cover plate of the retina display, and arranged the InI monitor for displaying the 3D image of the human model. Note that in this case we set the reference plane just behind the back of the model, so that all the body and mainly the left hand were reconstructed floating (about 3 cm the hand) in front of the monitor. This cannot be perceived in the video, but was clear for binocular observers. To show here the 3D nature of the displayed image we have recorded a video composed by views of the monitor obtained from different horizontal and vertical perspectives.

VI. Conclusion

In this paper, we have reported a novel procedure for the capture, processing and projection of integral images. Whereas the plenoptic camera is the best suited for the capture of integral images of small 3D scenes, the method proposed here can gain competitive advantage over other methods for the capture of integral images of big 3D scenes. Main advantage of fusing the kinect and the integral imaging concepts is the acquisition speed, and the small amount of handled data. Also the algorithms proposed are simple. We have demonstrated the utility of our method with two experiments, which show that full-parallax 3D images can be displayed by an InI monitor, and that calculated microimages can be adapted to the characteristics of the monitor. In further research we will combine the 3D points clouds obtained with a pair of Kinect, in order to tackle the problem of potential occlusions.

References

[1] M. Martínez-Corral and G. Saavedra, "The resolution challenge in 3D optical microscopy," *Progress in Optics*, vol. 52, pp. 1–67, 2009.

[2] M. Weber, M. Mickoleit, and J. Huisken, "Lightsheet microscopy," *Methods in Cell Biology*, vol. 123, pp. 193–215, 2014.

[3] J. Wang, H. Suenaga, K. Hoshi, L. Yang, E. Kobayashi, I. Sakuma, and H. Liao, "Augmented reality navigation with automatic marker-free image registration using 3-D image overlay for dental surgery," *IEEE Trans Biomed Eng.*, vol. 61, no. 4, pp. 1295–1304, Apr. 2014.

[4] M. Martinez-Corral, "Multiperpective Fundus camera," Spain Patent ES 2442178 B2, Jul. 23, 2014.

[5] J. Geng and J. Xie, "Review of 3-D endoscopic surface imaging techniques," *IEEE Sensors J.*, vol. 14, no. 4, pp. 945–960, Apr. 2014.

[6] I. Marcus, I. T. Tung, E. O. Dosunmu, W. Thiamthat, and S. F. Freedman, "Anterior segment photography in pediatric eyes using the Lytro light field handheld noncontact camera," *J. AAPOS*, vol. 17, pp. 572–577, 2013.

[7] U. Perwass and C. Perwass, "Digital imaging system, plenoptic optical device and data processing method," Germany Patent WO 2010121637 A1, Oct. 28, 2010.

[8] T. G. Georgiev and A. Lumsdaine, "Methods and apparatus for rich image capture with focused plenoptic cameras," U.S. Patent 8345144 B1, Jan. 1, 2013.

[9] C. D. Huston and C. Coleman, "System and method for creating an environment and for sharing a location based experience in an environment," U.S. 20130222369 A1, Aug. 29, 2013.

[10] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," *J. Phys.*, vol. 7, pp. 821–825, 1908.

[11] R. Ng, "Digital Light Field Photography," Ph.D. dissertation, Stanford Univ., Palo Alto, CA, USA, 2006.

[12] H. Arimoto and B. Javidi, "Integral three-dimensional imaging with digital reconstruction," *Opt. Lett.*, vol. 26, pp. 157–159, 2001.

[13] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," *ACM Trans. Graph.*, vol. 25, pp. 924–934, 2006.

[14] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with plenoptic camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, pp. 99–106, 1992.

[15] T. Georgiev and A. Lumsdaine, "The focused plenoptic camera and rendering," *J. Elect. Imag.*, vol. 19, no. 2, Feb. 2010.

[16] J. Geng, "Three-dimensional display technologies," *Adv. Opt. Photon.*, vol. 5, pp. 456–535, 2013.

[17] X. Xiao, B. Javidi, M. Martínez-Corral, and A. Stern, "Advances in three-dimensional integral imaging: Sensing, display, applications," *Appl. Opt.*, vol. 52, pp. 546–560, 2013.

[18] J. Hong, Y. Kim, H.-J. Choi, J. Hahn, J.-H. Park, H. Kim, S.-W. Min, N. Chen, and B. Lee, "Three-dimensional display technologies of recent interest: Principles, status, issues," *Appl. Opt.*, vol. 50, pp. H87–H115, 2011.

[19] M. Martinez-Corral, A. Dorado, H. Navarro, G. Saavedra, and B. Javidi, "3D display by smart pseudoscopic-to-orthoscopic conversion with tunable focus," *Appl. Opt.*, vol. 53, pp. E19–E26, 2014.

[20] "Lightfield based commercial digital still camera," [Online]. Available: http://www.lytro.com

[21] M. Miura, J. Arai, T. Mishina, M. Okui, and F. Okano, "Integral imaging system with enlarged horizontal viewing angle," *Proc. SPIE*, vol. 8384, p. 83840O, 2012.

[22] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar, "PiCam: An ultra-thin high performance monolithic camera array," *ACM Trans. Graphics*, vol. 32, no. 166, 2013.

[23] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multiview imaging and 3DTV," *IEEE Signal Processing Magazine*, vol. 24, pp. 10–21, 2007.

[24] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM SIGGRAPH 2005*.

[25] T. Nasrin, F. Yi, S. Das, and I. Moon, "Partially occluded object reconstruction using multiple Kinect sensors," *Proc. SPIE 9117*, p. 91171G, 2014.

[26] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, and R. Moore *et al.*, "Real-time human pose recognition in parts from single depth images," in *Proc. IEEE Computer Soc. Conf. on Computer Vision and Pattern Recogn. (CVPR)*, 2011.

[27] J. Kramer, N. Burrus, F. Echtler, H. C. Daniel, and M. Parker, "Multiple kinects," in *Hacking the Kinect, Apress*, 2012, pp. 207–246.

[28] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli, "Depth Mapping Using Projected Patterns," U.S. Patent 20080240502A1, Oct. 2, 2008.

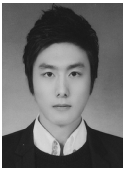
[29] M. Lee and J. Jeon, "Personal computer control using kinect," *Kor. Inst. Inf. Scientists and Eng.*, vol. 39, no. 1(A), 2012.

[30] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, vol. 12, pp. 1437–1454, 2012.

[31] M. Martínez-Corral, B. Javidi, R. Martínez-Cuenca, and G. Saavedra, "Formation of real, orthoscopic integral images by smart pixel mapping," *Opt. Express*, vol. 13, pp. 9175–9180, 2005.

[32] F. Okano, H. Hoshino, J. Arai, and I. Yayuma, "Real time pickup method for a three-dimensional image based on integral photography," *Appl. Opt.*, vol. 36, pp. 1598–1603, 1997.

[33] H. Navarro, R. Martínez-Cuenca, A. Molina-Martín, M. Martínez-Corral, G. Saavedra, and B. Javidi, "Method to remedy image degradations due to facet braiding in 3D integral imaging monitors," *J. Display Technol.*, vol. 6, no. 10, pp. 404–411, Oct. 2010.



Seokmin Hong received the B.Eng. and M.Sc. degrees in digital and visual contents from Dongseo University, Busan, Korea, in 2012 and 2014, respectively. In 2012, Dongseo University honored him with the B.Eng. Extraordinary Award.

Since 2012, he has been working with Institute of Ambient Intelligence, Dongseo University, Busan, Korea. His research interests are image processing, computer vision and applied computer science.

Donghak Shin received the B.S., M.S., and Ph.D. degrees in telecommunication and information engineering from Pukyong National University, Busan, Korea, in 1996, 1998, and 2001, respectively.

From 2001 to 2004, he was a senior researcher with TS-Photon established by Toyohashi University of Technology, Japan. From 2005 to 2006, he was with the 3D Display Research Center (3DRC-ITRC), Kwangwoon University, Korea. He worked as a research professor at Dongseo University in Korea from 2007 to 2010. He was a visiting scholar in electrical & computer engineering department at the University of Connecticut from 2011–2012. He is currently a senior research in Institute of Ambient Intelligence, Dongseo University in Korea. His research interests include 3D imaging, 3D displays, optical information processing, and holography.

Byung-Gook Lee received the B.S. degree in mathematics from Yonsei University, Korea, in 1987, and the M.S. and Ph.D. degrees in applied mathematics from Korea Advanced Institute of Science and Technology (KAIST), in 1989 and 1993, respectively.

He worked at the DACOM Corp. R&D Center as a senior engineer from March 1993 to February 1995. He has been working at Dongseo University since 1995. He is currently a Full Professor with the Division of Computer Information Engineering at Dongseo University, Korea. His research interests include computer graphics, computer aided geometric design, and image processing.

Adrián Dorado was born in Spain in 1988. He received the B.Sc. and M.Sc. degrees in physics from the University of Valencia, Spain, in 2011 and 2012, respectively.

Since 2010, he has been with the 3D Imaging and Display Laboratory, Optics Department, University of Valencia, Spain. His research interests include 3D imaging acquisition and display.

Genaro Saavedra received the B.Sc. and Ph.D. degrees in physics from Universitat de València, Spain, in 1990 and 1996, respectively. His Ph. D. work was honored with the Ph.D. Extraordinary Award.

He is currently Full Professor with Universitat de València, Spain. Since 1999, he has been working with the "3D Display and Imaging Laboratory", at the Optics Department. His current research interests are optical diffraction, integral imaging, 3D high-resolution optical microscopy and phase-space representation of scalar optical fields. He has published on these topics about 50 technical articles in major journals and 3 chapters in scientific books. He has published over 50 conference proceedings, including 10 invited presentations.

Manuel Martínez-Corral was born in Spain in 1962. He received the M.Sc. and Ph.D. degrees in physics from the University of Valencia, Spain, in 1988 and 1993, respectively. In 1993, the University of Valencia honored him with the Ph.D. Extraordinary Award.

He is currently Full Professor of Optics at the University of Valencia, where he is co-leader of the "3D Imaging and Display Laboratory".His research interest includes resolution procedures in 3D scanning microscopy, and 3D imaging and display technologies. He has supervised on these topics seven Ph. D. theses, two of them honored with the Ph.D. Extraordinary Award. He has published over eighty technical articles in major journals, and pronounced over thirty invited and five keynote presentations in international meetings. He has been member of the Scientific Committee in over twenty international meetings.

In 2010, Dr. Martinez-Corral was named Fellow of the SPIE. He is co-chair of the Three-Dimensional Imaging, Visualization, and Display Conference within the SPIE meeting in Defense, Security, and Sensing (Baltimore). He is Topical Editor of the IEEE/OSA JOURNAL OF DISPLAY TECHNOLOGY.

Paper II

Three-Dimensional Integral-Imaging Display From Calibrated and Depth-Hole Filtered Kinect Information

Seokmin Hong, Adrián Dorado, Genaro Saavedra, Juan Carlos Barreiro, and Manuel Martínez-Corral

in *Journal of Display Technology* 12, 1301-8 (2016).

Three-Dimensional Integral-Imaging Display From Calibrated and Depth-Hole Filtered Kinect Information

Seokmin Hong, Adrian Dorado, Genaro Saavedra, Juan Carlos Barreiro, and Manuel Martinez-Corral

***Abstract*—We exploit the Kinect capacity of picking up a dense depth map, to display static three-dimensional (3D) images with full parallax. This is done by using the IR and RGB camera of the Kinect. From the depth map and RGB information, we are able to obtain an integral image after projecting the information through a virtual pinhole array. The integral image is displayed on our integral-imaging monitor, which provides the observer with horizontal and vertical perspectives of big 3D scenes. But, due to the Kinect depth-acquisition procedure, many depthless regions appear in the captured depth map. These holes spread to the generated integral image, reducing its quality. To solve this drawback we propose here, both, an optimized camera calibration technique, and the use of an improved hole-filtering algorithm. To verify our method, we performed an experiment where we generated and displayed the integral image of a room size 3D scene.**

***Index Terms*—Bilateral filter, bilinear interpolation, camera calibration, integral imaging, kinect, median filter, 3D display.**

I. Introduction

CONVENTIONAL photography is fully adapted to record the 3D world scenes into a two-dimensional (2D) sensor. Although 2D images carry some cues from the 3D nature of scenes, they still lack important information. Fortunately, nowadays there are techniques that are able to record 3D information from 3D scenes. One interesting method is to record a depth map. A depth map can be obtained, for example, by the stereo vision technique, which takes profit from the disparity between the images captured with two cameras arranged horizontally [1], [2]. Other techniques are based on the projection of a random IR dot pattern [3], [4], or on time-of-flight technology [5]–[7]. Also interesting is to take profit from the vertical and horizontal views captured in with integral-imaging (InI) technology [8]–[12]. InI can provide 3D images in color with, quasi-continuous, horizontal and vertical parallax. For this reason it has been considered as one of the most promising technologies for next generation of 3D displays [13]–[18].

In a previous work we proposed the use of the depth map and a RGB image obtained with a Kinect to calculate an integral image and project it onto a 3D display system [14]. Although innovative, this research provided results that must be improved. The main problem of this previous research was the appearance of big holes in the depth map, which propagate up to the generated integral image. Another problem comes from the use the Kinect's software-development-kit (SDK). This method implements a mapping from IR camera to the RGB camera of the Kinect in order to merge the views of both cameras. But after applied, the SDK function produces some noise due to an error in decimal computation. In addition, this method can't make the mapping in the opposite direction.

In order to solve these problems, we propose some alternatives. First, we propose a new method for the camera calibration process between the two different sensors presented in the Kinect. This procedure is described in Section II. (Fig. 1(a) and (b)) Second, we propose to recover the lost depth information by using a filtering algorithm, which is explained in Section III (Fig. 1(c) and (d)). By applying these changes, we are able to obtain a better depth map, with lesser holes, due to the filtering and a better calibration process. With this improved 3D information, we generate higher quality of microimages. The microimages generation process is descripted in Section IV. Finally, in Sections V and VI, we provide experimental results and conclusions respectively (Fig. 1(e)).

II. Calibration Between the Two Different Type of Cameras of the Kinect

The Kinect device is well known for its capacity of capturing simultaneously, with two different types of camera, both, a color image and a dense depth map [19]. However, the field-of-view (FOV) of the two cameras are not matched properly. One solution to this drawback is the well known Camera Calibration Technique [20]–[25], which is able to correct the camera-lens distortions, to figure out the focal length and to estimate the 3D location of a camera in real world coordinate system. Furthermore, this process can determine the correlation between the camera's own coordinate's unit (image's pixel coordinate) and the real world's measurement unit (millimeters, centimeters, etc.).

The calibration is a two-step process. First, in order to find a relationship between two cameras (that is, in order to obtain their intrinsic and extrinsic parameters) a special pattern must be captured. Second, the two captures are merged together using

Manuscript received May 2, 2016; accepted July 21, 2016. Date of publication July 27, 2016; date of current version October 11, 2016. This work was supported in part by the Plan Nacional I+D+I under Grant DPI2015-66458-C2-1-R, Ministerio de Economía y Competitividad, Spain, by Generalitat Valenciana, Spain under Project PROMETEOII/2014/072. The work of S. Hong was supported by the University of Valencia under a predoctoral Grant UV-INV-PREDOC15-265754.

The authors are with the Department of Optics, University of Valencia, Burjassot E-46100, Spain (e-mail: seokmin.hong@uv.es; adrian.dorado@uv.es; Genaro.Saavedra@uv.es; barreiro@uv.es; manuel.martinez@uv.es).

This paper contains supplemental material available at http://ieeexplore.ieee.org.

Color versions of one or more of the figures are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JDT.2016.2594076

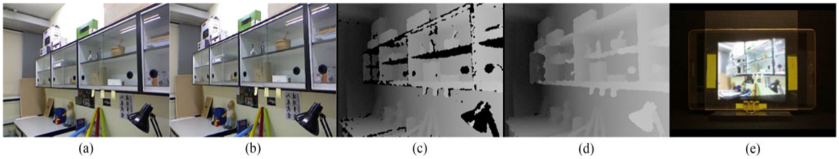


Fig. 1. From *left* to *right*: (a) Captured raw color intensity image. (b) Processed image: coordinate conversion from color intensity camera to Infra-Red (IR) camera of the Kinect by using calibrated camera parameters with bilinear interpolation method. (c) Captured raw depth map image. (d) Image obtained using our proposed depth-hole filtering algorithm. (e) Final result: single-frame excerpt from the recorded video of the implemented integral imaging monitor.

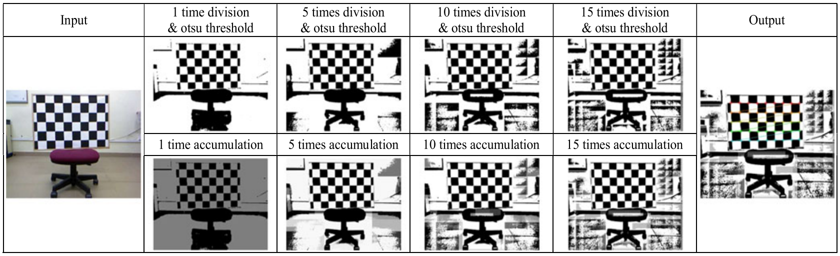

Input
1 time division & otsu threshold
5 times division & otsu threshold
10 times division & otsu threshold
15 times division & otsu threshold
Output
1 time accumulation
5 times accumulation
10 times accumulation
15 times accumulation

Fig. 2. Sequential steps of the process to detect the chessboard pattern: the image is segmented into different parts. Otsu threshold is applied to each part and the results are accumulated. The procedure depends on the scene; therefore, we applied several segmentations to the captured scene, from 1 to 15 times flexibly. Upper row shows the results of original Otsu algorithm. Bottom row shows the result of adding accumulative procedure.

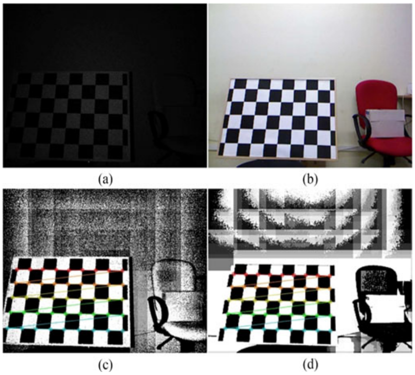


Fig. 3. Processed result using the proposed threshold technique. (a) Raw image from IR camera; (b) raw image from RGB camera; (c) thresholded image from (a); and (d) thresholded image from (b). Chessboard's corner points are found correctly even when the image has low illumination.

a transformation that takes into account the data previously obtained. Therefore, in order to calibrate we use a chessboard as the reference pattern for both cameras. The main reason of using a chessboard is that a regularized pattern improves the accuracy of the calibration [26].

It is worth to note that the image of the chessboard recorded with the IR camera is very dark, see for example Fig. 3(a) [27]. In order to overcome this drawback we propose to use a new algorithm that is based in the application of well-known Otsu thresholding method [28], but in iterative accumulative way. Our algorithm works as follows:

First, the image I is divided into i different parts, with a sequential ratio $S_i = 1/i1/i$. Then, Otsu algorithm, T, is applied to each individual part $I_i I$. Finally, the results are saved accumulatively into destination I_{dst}, as shown in Eqs. (1) and (2)

$$I_{\text{dst}} = \sum_{i=1}^{n} \left[\{T(I_i) + I_{\text{dst}}\}/2\right] \qquad (1)$$

where

$$I_i = S_i(I) \qquad (2)$$

The main feature of this new algorithm is that, independently of the complexity of the whole image, it highlights the chessboard pattern. Figs. 2 and 3 show this procedure in detail.

Once applied our algorithm, we can calculate the intrinsic and extrinsic parameters of the cameras. The intrinsic parameters are: the focal length, the aspect ratio and the central point of the view. The extrinsic parameters are the camera 3D location and orientation. We can use the values of theses parameters to fuse both cameras coordinate systems. The matrix Eqs. (3) to (5) rule the process to merge the FOV of the two cameras. In these equations K represents the intrinsic parameters; 2D point coordinate within each image is represented by p; P is each camera's

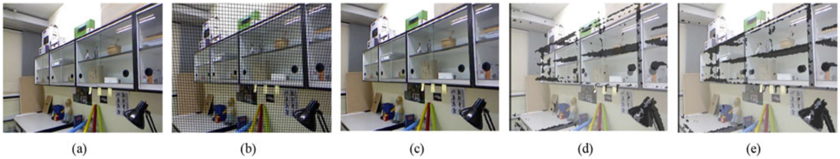


Fig. 4. From *left* to *right*: (a) Raw RGB intensity image; (b) calibrated RGB image from RGB camera to IR camera, (c) interpolated image's result and (d) and (e) raw mapping result in original and calibrated scene respectively. We can see that there are gaps (like as black stripes) in image (b). This is due to an error coming from the decimal computation procedures between each calibrated pixels. We can fix this by using a bilinear interpolation; the result of the interpolation is shown in (c). To attest this calibrated result, we make a mapping between the RGB image and depth map image. From the panel (e), we can check that the calibrated RGB image fuses to depth map image well.

coordinate 3D point. Finally R and T are rotation and translation matrices. These equations also permit to do an inverse mapping, so that it is possible to make the mapping (merging the FOV) between the two cameras in both directions

$$P_{\mathrm{rgb}} = \mathrm{inv}\,(K_{\mathrm{rgb}}) \times p_{\mathrm{rgb}} \quad (3)$$

$$P_{\mathrm{ir}} = R \times P_{\mathrm{rgb}} + T \quad (4)$$

$$p_{\mathrm{ir}} = K_{\mathrm{ir}} \times P_{\mathrm{ir}} \quad (5)$$

Unfortunately, there is still another problem related with the calibration process. The position of a pixel in an image is given in natural numbers. The final calibrated pixel coordinate (the one with the merged FOV) is represented by real numbers, which are rounded. This generates some gaps within calibrated pixels, see Fig. 4(b). As result, many pixels are misaligned in the calibrated image and therefore, into the 3D point cloud also same. To solve this problem, we applied a bilinear interpolation to the empty pixels between calibrated pixels; see Fig. 4 (c).

Finally, taking this into account, the calibrated RGB image can be mapped well to the IR image. Also, we are able to display in real time the calibrated images of both cameras. After all this calibration process, now it is possible to use the depth map and the RGB image to generate a 3D virtual point cloud in which we assign to each point its corresponding 3D position and RGB intensity. The Fig. 5 shows two views of a single shot of the virtual 3D point cloud corresponding to the recorded scene.

III. Depth Hole Filtering

From the information captured with the Kinect we can compose a collection of microimages ready to be displayed on an InI monitor, as we showed in our previous paper [14]. However, that research had an important drawback. There were depthless pixels in the recorded depth map, which generated noise into the calculated microimages.

In the Kinect, the IR light source emits a known pattern and the depth information is calculated after comparison, by using triangulation method, between the known illumination pattern and the observed dots at the captured scene [29]. The problem arises when some reflective surfaces reflect IR light into another direction or when the IR light penetrates into transparent surfaces. This produces a loss in the depth information provided by the Kinect and thus, generates the holes in the depth map.

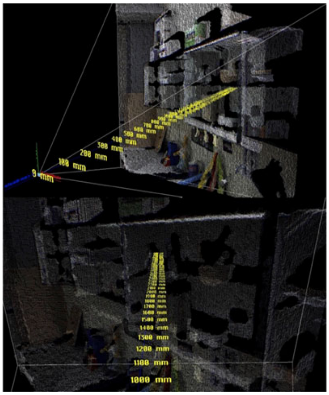


Fig. 5. 3D point cloud in the virtual space. From both panels, it is clear that each point has its own (x, y, z) position and RGB color intensity. Note that each 3D point is ordered on the basis on real world's measurement unit.

To avoid this drawback, we propose here the use of a depth-hole filtering process based on Camplani and Salgado work [30]. In order to make the algorithm more efficient, we propose here some improvements on the original version of the algorithm. The key idea of Camplani and Salgado filtering process is iteration. In their proposal the depth map is captured several times. Every acquired depth-map frame is filtered in order to remove the spatial noise and purify the object boundaries. This filtered depth map is used to update both, the depth model and the filtering algorithm. Therefore, each acquired depth map increases the quality of the depth model and the applied filter. So, after any iteration more reliable depth information is obtained.

The flow chart of the hole-filtering algorithm, including our proposed improvement, is shown in Fig. 7. The real depth information D and the color intensity I are captured at every loop. A computed depth-map model D_{model} and consistent depth map C_{depth} are the core of this algorithm. The D_{model} is the result of applying the filter to the depth map and the C_{depth} is a version of the depth map that only stores the maximum depth values of all the iterations results.

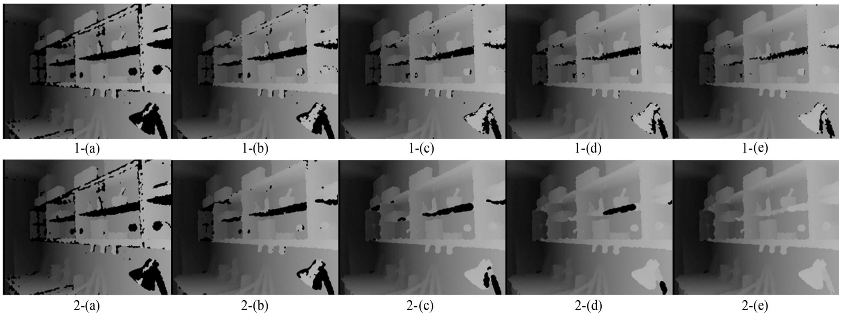


Fig. 6. Comparison between original (1-(a)–(e), see also Media 1) and proposed (2-(a)–(e), see also Media 2) filtering algorithm: (a) is initial frame, (b) 5 iterations, (c) 20 iterations, (d) 80 iterations, and (e) 219 iterations. Through the panels, we can compare the process clearly. Above all things, the proposed strategy can recover the depthless pixel more efficient than the original method.

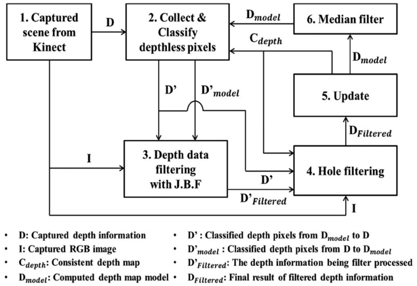

1. Captured scene from Kinect
D
2. Collect & Classify depthless pixels
D_model
C_depth
6. Median filter
D_model
5. Update
D'
D'_model
D_Filtered
I
3. Depth data filtering with J.B.F
D'
4. Hole filtering
D'_Filtered
I
• D: Captured depth information
• I: Captured RGB image
• C_depth: Consistent depth map
• D_model: Computed depth map model
• D' : Classified depth pixels from D_model to D
• D'_model : Classified depth pixels from D to D_model
• D'_Filtered: The depth information being filter processed
• D_Filtered: Final result of filtered depth information

Fig. 7. Flow chart of the proposed hole filtering strategy.

After capturing the 3D scene information, the second step is collecting and classifying the depthless pixels by using the captured depth information and a computed depth-map model. If D has depthless pixels, they are replaced by the corresponding pixel of D_{model} if the pixel value is reliable (if C_{depth} is greater than threshold value d_{thres}). And if D_{model} has depthless pixels, they are replaced by the corresponding pixel from D. Due to this change in information, D and D_{model} become D' and D'_{model} respectively. Note that on the first iteration, the value of D_{model} and C_{depth} are 0, and D' will be assigned with the value of D.

Next, the depth data is filtered using a joint (or cross) bilateral filter (JBF) [31], in order to improve the classified depth information's accuracy. JBF makes depth values reliable and is able to distinguish edges from surface's regions by checking and comparing neighbor pixels on both, the depth map and the RGB image. Note that JBF is an improved version of the similarity kernel of the bilateral filtering technique. The bilateral filtering is an edge-preserved and noise-reduced smoothing filter. To manage each pixel the filter has only two main kernel functions: the similarity kernel and the closeness kernel. These kernels are based on a Gaussian distribution and the pixel value is replaced by a weighted-average from their neighbor pixels [32].

The JBF works as follows; $c(j,k)$ is the domain term like as bilateral filter, $s(\|D'^{j}_{\mathrm{model}} - D'^{k}_{\mathrm{model}}\|)$ is the similarity kernel in classified D_{model} and $s(\|I^j - I^k\|)$ is from the similarity kernel of color intensity. The scalar R^j is a normalization factor, and all of its calculated result is represented by D'_{filtered} [see, the Eqs. (6) and (7)]

$$D'^{j}_{\mathrm{filtered}} = 1/R^j \iint_{k\in\Omega j} D'^{k} c(j,k)\, s\left(\left\|D'^{j}_{\mathrm{model}} - D'^{k}_{\mathrm{model}}\right\|\right) s\left(\left\|I^j - I^k\right\|\right) \tag{6}$$

where

$$R^j = \iint_{k\in\Omega j} c(j,k)\, s\left(\left\|D'^{j}_{\mathrm{model}} - D'^{k}_{\mathrm{model}}\right\|\right) s\left(\left\|I^j - I^k\right\|\right) \tag{7}$$

The fourth step consists on improving the previous filtered result. If D'_{filtered} still has some depthless pixels or regions, some of the missing depth information can still be recovered using all the data previously obtained. $H(C_{\mathrm{depth},}\ \Omega_j)$ is a binary function that evaluates which pixels need to be updated with the information stored in D' and I. $c(j,k)$ and $s(I^j - I^k)$ are the same filtering functions as Eq. (6)

$$D^{j}_{\mathrm{filtered}} = H\left(C_{\mathrm{depth}}, \Omega_j\right)/R^j \iint_{k\in\Omega j} D'^{k} c(j,k)\, s\left(\left\|I^j - I^k\right\|\right) \tag{8}$$

where

$$R^j = \iint_{k\in\Omega j} D'^{k} c(j,k)\, s\left(\left\|I^j - I^k\right\|\right) \tag{9}$$

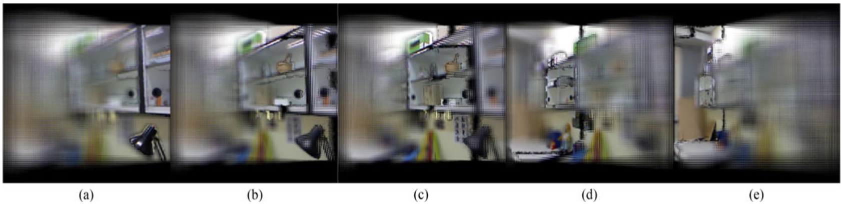


Fig. 8. Collection of microimages generated from the 3D points cloud captured with the Kinect. These panels show about different focused planes; (a) is focused in 870 mm, (b) is 1290 mm, (c) is 1520 mm, (d) is 2385 mm, (e) is 3145 mm, respectively. Each focused plane shows an object with clear shape.

$$H\left(C_{\mathrm{depth}},\Omega_j\right) = \begin{cases} 1 \text{ if count}\left[C_{\mathrm{depth}}\ \left(\Omega_j\right) > d_{\mathrm{thres}}\right]/\text{Area}\left(\Omega_j\right) > th_{\%} \\ \quad\quad 0 \text{ otherwise} \end{cases}. \tag{10}$$

The fifth's step is to update the filtered depthless pixels into both, C_{depth} and D_{model}. Parameter α is a constant weight factor whose value is obtained from our empirical evidence. The aim of this value is to obtain stability on the process, giving more importance to the previous results

$$D^j_{\mathrm{model}} = \alpha D^j_{\mathrm{filtered}} + (1-\alpha)\, D^j_{\mathrm{model\_OLD}} \tag{11}$$

$$C^j_{\mathrm{depth}} = \begin{cases} D^j_{\mathrm{model}}\ if\ D^j_{\mathrm{model}} > C^j_{\mathrm{depth}} \\ \quad \text{otherwise} \end{cases}. \tag{12}$$

The last step is the application of a median filter, which is our contribution to the process. It helps to expand reliable depth values into their neighbor pixels or clean up the noise in object's boundary/edge regions. The filter chooses the medium value between its neighbor pixels. For that reason, it can remove efficiently and correctly small rubbish particles and, as result, D_{model} and C_{depth} are updated and becomes a reliable filtered and computed result.

After a few repetitions using this proposed filtering process, we can get a clear hole-filled depth map. Fig. 6 shows the results of applying both, the original algorithm and the proposed one, to some specific frames. All the images in Fig. 6 correspond to the consistent depth map C_{depth}. Note that when we add the median filter, the small and big depth-hole regions were recovered more efficiently than in the original algorithm [30]. Also, instead of the traditional 256 depth scales used in the original paper we used a real depth scale of 3976 (chosen for empirical reasons). This means that we have more abundant depth information than the original one.

To finish this section we summarize the parameters used in the algorithm in Table I.

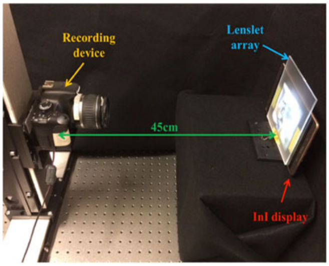


Fig. 9. The overview of our experimental system. We moved the recording device vertically and horizontally to record different perspectives of the integrated image.

TABLE I
ALGORITHM PARAMETERS

δ for closeness filter c	4.5
δ for similarity filter s for the depth (0 – 3975 scales)	9×9
δ for similarity filter s for the color (0 – 255 scales)	9×9
d_{thres}	5
$th_{\%}$	0.65
α	0.04
Median filter size	7×7

IV. MICROIMAGES GENERATION

In order to generate the microimages, we follow a process equivalent to the one reported in our previous paper [14], but adapted to a new display device. Specifically, in our experiment the InI monitor is composed by a Samsung SM-T700 (359 pixels/inch), and a micro-lenslet-array (MLA) consisting of 113 $\times$ 113 lenslets of focal length $f_L = 3.3$ mm and pitch $p = 1.0$ mm (Model 630 from Fresnel Technology). The generated microimages are then composed by 15×15 px, the gap between the microlenses and the display is fixed to $g = 49.5$ px, and the full size of the integral image is 1695×1695 px.

The VPA used to capture the synthetic microimages is placed into the point cloud's coordinate system. Note that the position of the VPA will determine the reference plane. Then we project, using projection mapping, each point of the 3D cloud through each pinhole of the VPA to obtain the microimages, as in [33]. We resize the image to 1597×1197 px to take into account the resolution of the display system $(14.13\ \mathrm{px/mm})$.

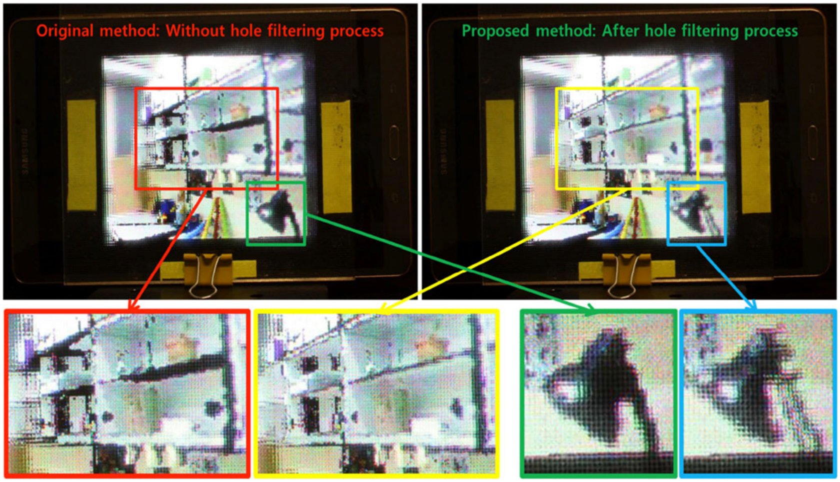

Original method: Without hole filtering process
Proposed method: After hole filtering process

Fig. 10. Comparison between the original method (*Top-Left* image, and also Media 3) and the proposed method (*Top-Right* image, and also Media 4). We have highlighted some specific parts of the images. It is clear that the result obtained with our proposed method shows more abundant 3D information.

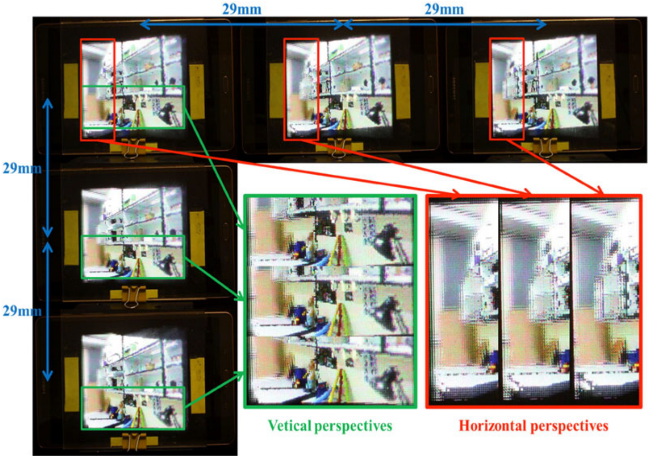


Fig. 11. Different views of the InI monitor in the vertical and horizontal direction, showing that the InI monitor has full parallax. The distance between different images is 29 mm. The total viewing zone is of 58 × 58 mm. A video of the views is shown in Media 4.

In particular we show, in Fig. 8, the microimages calculated from a VPA situated at different positions. The reference plane position determines which parts of the 3D image are in front or behind the screen.

V. Displayed 3D Image

Finally, the generated microimages are displayed onto our InI monitor. The MLA was properly aligned in front of the display system. The InI monitor displays and integrates the microimages towards the observer's eyes. Thus, a binocular observer can see some parts of the displayed scene in front of the monitor and some other behind. However this full-parallax effect cannot be directly observed in a manuscript or even in a video. In order to demonstrate here this effect we proceeded as follows. First we replaced the observer by a monocular digital camera. The Fig. 9 shows our experimental system's overview. Then we obtained a collection of pictures after displacing horizontally and vertically the camera along a region of 58×58 mm. With these pictures we composed a video in which the InI monitor was observed from different horizontal and vertical perspectives. The Figs. 10 and 11 show the experimental results with more clarity. As you can see in Fig. 10, the hole-filtered depth map generates better images, recovering some of the lost depth information in the original one. In Fig. 10, we have highlighted with color rectangles the areas where the differences are clearly shown. Finally, Fig. 11 shows the different perspectives, vertical and horizontal, of the InI display system.

VI. Conclusion

In this paper, we have reported how to generate improved microimages using manipulated 3D information, obtained with a Kinect device. For that, we use the camera calibration technique with bilinear interpolation method. Also, we have proposed an efficient hole-filtering algorithm to fill the depth holes, which appear in the depth map captured by the Kinect. Therefore, this well-refined depth information reduces the noise in the recorded 3D information. In order to project our synthesized 3D information onto an InI display system, we generate microimages by using projection mapping through a VPA. To demonstrate the utility of our proposal, we projected the microimages onto an InI monitor, providing different, depth-hole free and continuous horizontal and vertical perspectives to the observer.

References

[1] Bumblebee2, (2006, Aug. 23). [Online]. Available: http://www.ptgrey.com/stereo-vision-cameras-systems

[2] ZED, (2015, Feb. 18). [Online]. Available: https://www.stereolabs.com/

[3] Microsoft Kinect v1.0, (2010, Nov. 4). [Online]. Available: http://www.xbox.com/en-US/xbox-360/accessories/kinect

[4] Primesense Calmine 1.08 & 1.09, (2013). [Online]. Available: https://en.wikipedia.org/wiki/PrimeSense

[5] Canesta sensor, (2002). [Online]. Available: https://en.wikipedia.org/wiki/Canesta

[6] Microsoft Kinect v2.0, (2013, Nov. 22). [Online]. Available: http://www.xbox.com/en-US/xbox-one/accessories/kinect-for-xbox-one

[7] Intel Realsense, (2014, Jan. 6). [Online]. Available: https://software.intel.com/en-us/realsense/

[8] X. Xiao, B. Javidi, M. Martínez-Corral, and A. Stern, "Advances in three-dimensional integral imaging: Sensing, display, and applications," *Appl. Opt.*, vol. 52, pp. 546–560, 2013.

[9] Lytro camera, (2011, Oct. 19). [Online]. Available: https://www.lytro.com/

[10] PiCam: Pelican Imaging Camera, (Nov. 2013) [Online]. Available: http://www.pelicanimaging.com/

[11] Raytrix camera, (2010). [Online]. Available: http://www.raytrix.de/

[12] B. Javidi, J. Sola-Pikabea, and M. Martínez-Corral, "Breakthroughs in photonics 2014: Recent advances in 3D integral imaging sensing and display," *IEEE Photon. J.*, vol. 7, 2015, Art. no. 0700907.

[13] S. Hong, D. Shin, J. Lee, and B. Lee, "Viewing angle-improved 3D integral imaging display with eye tracking sensor," *J. Inf. Commun. Convergence Eng.*, vol. 12, pp. 208–214, 2014.

[14] S. Hong *et al.*, "Towards 3D television through fusion of kinect and integral-imaging concepts," *J. Display Technol.*, vol. 11 no. 11, pp. 894–899, Nov. 2015.

[15] A. Stern and B. Javidi, "Three-dimensional image sensing, visualization, and processing using integral imaging," *Proc. IEEE*, vol. 94, no. 3, pp. 591–607, Mar. 2006.

[16] F. Okano, H. Hoshino, J. Arai, and I. Yayuma, "Real time pickup method for a three-dimensional image based on integral photography," *Appl. Opt.*, vol. 36, pp. 1598–1603, 1997.

[17] J. S. Jang and B. Javidi, "Improved viewing resolution of three-dimensional integral imaging by use of nonstationary micro-optics," *Opt. Lett.*, vol. 27, no. 5, pp. 324–326, 2002.

[18] Y. Kim, K. Hong, and B. Lee, "Recent researches based on integral imaging display method," *3D Res.*, vol. 1, no. 1, pp. 17–27, 2010.

[19] Specification of Kinect v1.0, (2013, Sep. 17). [Online]. Available: http://msdn.microsoft.com/en-us/library/jj131033.aspx

[20] Camera calibration in OpenCV, (2000, Jun.). [Online]. Available: http://docs.opencv.org/doc/tutorials/calib3d/camera_calibration/camera_calibration.html

[21] Camera calibration toolbox from Caltech, (2004, Jul. 20). [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/

[22] Robust multi-camera calibration, (2005, Jun. 7). [Online]. Available: http://graphics.stanford.edu/~vaibhav/projects/calib-cs205/cs205.html

[23] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proc. IEEE 7th Int. Conf. Comput. Vis.*, 1999, pp. 666–673.

[24] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[25] P. Sturm and S. Maybank, "On plane-based camera calibration: A general algorithm, singularities, applications," presented at the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Fort Collins, CO, USA, 1999, pp. 432–437.

[26] B. Gary and A. Kaehler, *Learning OpenCV: Computer Vision With the OpenCV Library*. North Sebastopol, CA, USA: O' Reilly Media Press, 2008.

[27] Infrared stream in Kinect v1.0, (2013, Sep.). [Online]. Available: http://msdn.microsoft.com/en-us/library/jj663793.aspx

[28] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Sys., Man., Cyber.*, vol. 9, no. 1, pp. 62–66, Jan. 1979. doi:10.1109/TSMC1979.4310076

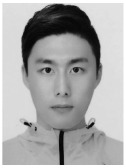
[29] D. Alex Butler *et al.*, "Shake'n'sense: Reducing interference for overlapping structured light depth cameras," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2012, pp. 1933–1936.

[30] M. Camplani and L. Salgado, "Efficient spatio-temporal hole filling strategy for Kinect depth maps," in *Proc. SPIE Int. Conf. 3-D Image Process. Appl.*, 2012, pp. 1–10.

[31] G. Petschnigg *et al.*, "Digital photography with flash and no-flash image pairs," in *Proc. Annu. Conf. Comput. Graph. Interactive Techn.*, 2004, pp. 664–672.

[32] C. Tomasi, and R. Manduchi, "Bilateral filtering for gray and color images,"in *Proc. 6th Int. Conf. Comput. Vis.*, 1998, pp. 839–846.

[33] M. Martínez-Corral, B. Javidi, R. Martínez-Cuenca, and G. Saavedra, "Formation of real, orthoscopic integral images by smart pixel mapping," *Opt. Express*, vol. 13, pp. 9175–9180, 2005.



Seokmin Hong received the B.Eng. and M.Sc. degrees in digital and visual contents from Dongseo University, Busan, Korea, in 2012 and 2014, respectively.

From 2012 to 2014 he was working with the Institute of Ambient Intelligence Laboratory, Dongseo University, Korea. Since 2015 he has been working with the 3D Imaging and Display Laboratory, Optics Department, University of Valencia, Spain. His research interests include image processing, computer vision, and applied computer science.

Mr. Hong received the B.Eng. Extraordinary Award from Dongseo University in 2012.

Adrian Dorado was born in Spain in 1988. He received the B.Sc. and M.Sc. degrees in physics from the University of Valencia, Spain, in 2011 and 2012, respectively. Since 2010, he has been with the 3D Imaging and Display Laboratory, Optics Department, University of Valencia, Spain. His research interests include 3D imaging acquisition and display.

Genaro Saavedra received the B.Sc. and Ph.D. degrees in physics from the Universitat de València, Spain, in 1990 and 1996, respectively.

He is currently a Full Professor of optics at Universitat de València, Spain, where he is Coleader of the "3D Imaging and Display Laboratory." His current research interests include optical diffraction, integral imaging, 3D high-resolution optical microscopy, and phase-space representation of scalar optical fields. He has published on these topics about 70 technical articles in major journals and three chapters in scientific books. He has published over 50 conference proceedings, including 10 invited presentations.

Dr. Saavedra received the Ph.D. Extraordinary Award for his Ph.D. work.

Juan Carlos Barreiro received the B.Sc. (honored with an Extraordinary Award) and Ph.D. degrees in physics from the Universitat de València, Valencia, Spain, in 1985 and 1992, respectively.

He is currently an Associate Professor at the Department of Optics, Universitat de València. In addition, he has carried out research at CIEMAT, Madrid, Spain, and as a Postdoctoral Research Fellow at INAOE, Puebla, México and at The Institute of Optics, University of Rochester, Rochester, NY, USA. His current research interests include optical diffraction, high-resolution optical microscopy, and integral imaging. Besides that, he is actively involved in promoting Physics and Optics Education through outreach activities.

Manuel Martinez-Corral was born in Spain in 1962. He received the M.Sc. and Ph.D. degrees in physics from the University of Valencia, Valencia, Spain, in 1988 and 1993, respectively.

He is currently a Full Professor of optics at the University of Valencia, where he is Coleader of the "3D Imaging and Display Laboratory." His research interests include resolution procedures in 3D scanning microscopy, and 3D imaging and display technologies. He has published over hundred technical articles in major journals, and pronounced more than 40 invited and five keynote presentations in international meetings. He is a Topical Editor of the IEEE/OSA JOURNAL OF DISPLAY TECHNOLOGY.

Dr. Martinez-Corral has supervised 12 Ph.D. theses, three of them received the Ph.D. Extraordinary Award. In 1993 the University of Valencia honored him with the Ph.D. Extraordinary Award. He has been a Member of the Scientific Committee for more than 20 international meetings. He is Cochair of the Three-Dimensional Imaging, Visualization, and Display Conference within the SPIE meeting in Defense, Security, and Sensing (Baltimore). In 2010 he was named Fellow of the SPIE.

Paper III

Full parallax three-dimensional display from Kinect v1 and v2

Seokmin Hong, Genaro Saavedra, and Manuel Martínez-Corral

in *Optical Engineering* 56, 041305 (2017).

Paper III

Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Full parallax three-dimensional display from Kinect v1 and v2

Seokmin Hong
Genaro Saavedra
Manuel Martinez-Corral

SPIE. Seokmin Hong, Genaro Saavedra, Manuel Martinez-Corral, “Full parallax three-dimensional display from Kinect v1 and v2,” *Opt. Eng.* **56**(4), 041305 (2017), doi: 10.1117/1.OE.56.4.041305.

Optical Engineering 56(4), 041305 (April 2017)

Full parallax three-dimensional display from Kinect v1 and v2

Seokmin Hong,* Genaro Saavedra, and Manuel Martinez-Corral
University of Valencia, 3-D Imaging and Display Laboratory, Department of Optics, Burjassot, E-46100, Spain

Abstract. We exploit the two different versions of Kinect, v1 and v2, for the calculation of microimages projected onto integral-imaging displays. Our approach is based on composing a three-dimensional (3-D) point cloud from a captured depth map and RGB information. These fused 3-D maps permit to generate an integral image after projecting the information through a virtual pinhole array. In our analysis, we take into account that each of the Kinect devices has its own inherent capacities and individualities. We illustrate our analysis with some imaging experiments, provide the distinctive differences between the two Kinect devices, and finally conclude that Kinect v2 allows the display of 3-D images with very good resolution and with full parallax. *© 2016 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.OE.56.4.041305]

Keywords: three-dimensional display; integral imaging; point cloud; Kinect v1; Kinect v2.

Paper 161256SS received Aug. 9, 2016; accepted for publication Sep. 28, 2016; published online Oct. 19, 2016.

1 Introduction

Recently, integral imaging (InI) has been considered as one of the potential technologies in order to display real world scenes. Conventionally, the pickup stage of InI is performed by inserting a tiny lens array in front of a two-dimensional (2-D) imaging sensor. A remarkable feature of the InI technique is that every captured picture involves different perspectives information. The reason is that optical rays proceeding from three-dimensional (3-D) objects are collected by every lens, and recorded by the imaging sensor with different incidence angles. Here, we name as microimage the image recorded behind any microlens. The whole array of microimages is named here as the integral picture. When the integral picture is projected onto an InI monitor, it can provide the observers with 3-D floating color images, which have full-parallax and quasicontinuous perspective.[1–4] Many researchers have applied the InI technique in different fields.[5–15]

Meanwhile, there are various depth-sensing 3-D imaging techniques announced to record 3-D scenes. Among them, one interesting technique is stereovision, which exploits the disparity information from two arranged cameras.[16,17] However, in the past few years, the use of technologies related to infrared (IR) light sensors[18–21] has become increasingly popular. Especially the Kinect device from Microsoft that profits from IR lighting technology in the case of depth acquisition. Until now, there are two different versions of Kinect. The main commercial specifications of them are described in Table 1. The Kinect allows acquiring RGB, IR, and depth maps in real-time with a high frame rate. For that reason, many researchers are now interested in its capability. As is well known, both sensors have many different features for obtaining a dense depth map. The Kinect v1 uses a structured IR dot-pattern emitter and IR camera to evaluate depth information. In comparison, the Kinect v2 utilizes time-of-flight (ToF) technology, which consists of emitting IR flashes at high frequency. Having IR light that reflects from most 3-D surfaces, the sensor can evaluate the depth distance by measuring the light's returning time.[22,23] The main drawback of both, Kinect v1 and Kinect v2, is that they are limited for long range. Comparable results, but with an extended range, has been demonstrated but with a different technology.[24]

2 Calibration of Kinect v1 and v2

As seen in Table 1, the commercial specifications of the Kinects do not reflect all the characteristics of those devices. In order to extend this information, and also to confirm some commercial parameters, we performed a number of experiments.

2.1 *Coupled Area at the Scene*

The aim of our first experiment was to find the common area in the scene, and to check both the RGB and IR camera's fields of view (FOV) through empirical parameters. For this experiment, we first defined the standpoint of Kinect devices as the position of the nut where the tripod is screwed in. Then we defined an optical axis and set the Kinect frontal face parallel to the target. As the common target, we choose a chessboard pattern, which has simple and repetitive shapes and permits to easily detect feature points. Most of all, the regularized pattern influence improves the accuracy of the calibration's result.[25] We find common correspondence features in each captured scene and calculate correlation parameters, which are called homography, projectivity, or projective transformation. These parameters represent a general plane-to-plane correlation equation in a projective plane. These values convince to map from one camera's 2-D view to another.[26–29] Figure 1 shows the common area in both Kinect devices.

*Address all correspondence to: Seokmin Hong, E-mail: seokmin.hong@uv.es

0091-3286/2016/$25.00 © 2016 SPIE

Table 1 Comparison between Kinect v1 and v2 specifications.

List	Kinect v1	Kinect v2
Released (year)	2010	2014
RGB camera (pixel)	640 × 480 (max: 1280 × 960)	1920 × 1080
FPS in RGB camera	30 (Max: 12)	30 (low-light: 15)
IR camera (pixel)	640 × 480	512 × 424
FPS in IR camera	30	30
Depth acquisition method	Structured IR light pattern	ToF
Depth distance (mm)	800–4000	500–4500
Horizontal FOV (deg)	57	70
Vertical FOV (deg)	43	60

2.2 *Comparison of Field of View with Empirical Parameters*

Next, we attempt to measure, for both Kinects, the RGB and IR FOVs. Actually, the official specification did not mention the RGB FOV. For that reason, we measured the FOVs with two methods: (a) estimate FOV by calibrated camera parameters and (b) physical calibration progress at a certain distance. First, we use the calibrated camera parameters reported in Ref. 22 and calculate each FOV by using Eq. (1). In this equation, C_w and C_h are the RGB and IR physical imaging sensor sizes, and f is a focal length. R_w and R_h are the calculated angles in the horizontal and vertical directions.[30] Finally, we derived FOVs from the referenced parameters (see Table 2).

$$R_w = 2\arctan\left(\frac{C_w}{2f}\right), \quad R_h = 2\arctan\left(\frac{C_h}{2f}\right). \tag{1}$$

Second, we set up an environment in order to measure the FOV in a physical calibration progress. We placed the camera in a perpendicular direction from the wall within a certain distance. Then we stitched a piece of retroreflective (RR) sheet to the wall in the border area of the captured scene. Here, the IR camera can only capture an IR light and discard other light sources. Again, the IR camera cannot detect diffusing surfaces that are normal to the optical axis. In contrast, RR sheets can directly reflect IR light to the camera and, as a result, provide an easy way to verify a target's position in the IR camera's scene. From now, we already know about the Z-axis and width distance in millimeters. Then we can derive both horizontal and vertical FOVs using trigonometric function calculations. We illustrate this progress in Fig. 2 and put our empirical results in Table 3.

One thing worth noting is that through our experiment we confirmed some important issues. First, the two types of FOVs do not map properly. Some regions overlap, but not all parts from the scene are covered. Second, the announced FOV information from commercial specifications is for the IR camera. Third, we have proven that the commercial parameters are reliable. Parameters reported in Ref. 22 are also reliable, but not in the case of the FOV of the IR camera of Kinect v1.

3 Microimages Generation from Three-Dimensional Point Cloud

The aim of this research is to analyze and compare the two Kinect devices when they are concentrated in a specific

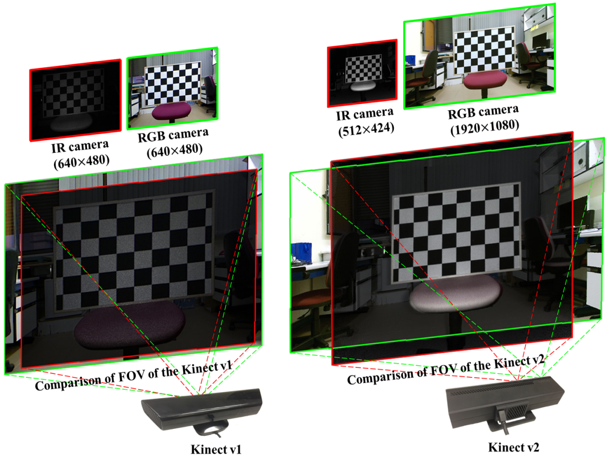


Fig. 1 Kinect v1 and v2 overlapped region in the captured scene. Green rectangle represents the RGB view and the red rectangle is the IR view.

Table 2 FOV result from calibrated camera parameters.

List	Kinect v1		Kinect v2	
Camera type	RGB	IR	RGB	IR
Focal length (mm)	3.099	6.497	3.291	3.657
Imaging sensor: width size (mm)	3.58	6.66	6.0	5.12
Imaging sensor: height size (mm)	2.87	5.32	3.38	4.24
Calculated FOV: horizontal (deg)	60.02	54.27	84.70	69.99
Calculated FOV: vertical (deg)	49.69	44.53	54.36	60.20

Table 3 Kinect v1 and v2's RGB, IR camera's FOV calculation result from physical calibrating progress.

List	Kinect v1		Kinect v2	
Camera type	RGB	IR	RGB	IR
Width distance (mm)	1177	1101	1814	1394
Height distance (mm)	912	816	1029	1143
Z-axis distance (mm)	1000	1000	1000	1000
Calculated FOV: horizontal (deg)	60.95	57.67	84.42	69.75
Calculated FOV: vertical (deg)	49.03	44.39	54.45	59.50

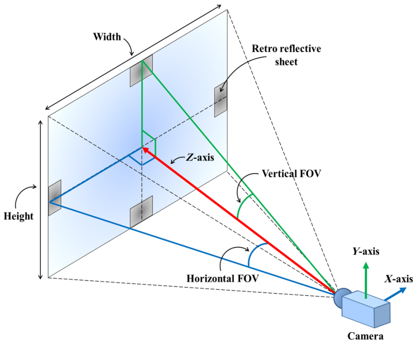


Fig. 2 The overview of our manipulated system. We put the camera at a certain distance from the wall and measure both vertical and horizontal distances.

application: the calculation of the collection of microimages that are projected onto an InI monitor with an aim of displaying 3-D images with full-parallax.

The procedure for calculation of microimages is as follows. First, the captured RGB and depth map images (see Fig. 3) are modified into a 3-D point cloud, following Ref. 6. From this result (see Fig. 4), we confirmed that Kinect v2 is able to capture the depth information of further points. Moreover, the density of point cloud data is also different. The Kinect v1, for instance, has a specific layer structure [see Fig. 4(a)]. But Kinect v2 provides more dense depth information without any regularized figuration [see Fig. 4(b)]. The most impressive feature from Kinect v2 is that this device can acquire depth information of slender targets, reflective surfaces, or even transparent objects compared with Kinect v1. In the third step of the procedure, we placed a virtual pinhole array (VPA) at a certain distance from the point cloud.

An important thing is that the VPA position decides the front and rear volumes in the displayed 3-D scenes. Due to this, the VPA position defines what we call the "reference plane" of the 3-D scene. In this experiment, we placed the VPA just behind the second chair. We assigned each 3-D point into microimages by back projection through the pinholes, as in Ref. 31. The main issue is that different features of the 3-D point clouds fully reflect into generated microimages. It is important to point out that the calculation of the microimages needs to take into account the parameters of the InI display. Specifically, we need to know the number of microlenses, their pitch, the gap, and the number of pixels behind any microlens. Figure 5 shows the calculated microimages, which are ready for projection into the InI display system described below. These two figures clearly show the differences of the two devices.

4 Experimental Results of Displayed Three-Dimensional Image

In order to display our microimages, we used the Samsung tablet SM-T700 (359 pixels/inch) as a high definition display, and a microlens array (MLA) consisting of 113 × 113 lenslets of focal length $f_L = 3.3$ mm and pitch $p = 1.0$ mm (Model 630 from Fresnel Technology). The resulting microimages are composed of 15 pixels. The gap between the microlenses and the display was fixed to $g = 49.5$ px. Finally, the full size of integral picture is 1695 × 1695 pixels. After fixing and aligning the MLA with the tablet, the resulting InI monitor displayed 3-D images with full parallax.

To demonstrate the three-dimensionality of the displayed images, we implemented the setup shown in Fig. 6, and recorded pictures of the InI display from many vertical and horizontal perspectives. From the pictures' collection, we composed two videos, one for the Kinect v1 (Video 1) and the other for the Kinect v2 (Video 2). Additionally, we excerpted a pair of frames from any video. These frames are shown and compared in Fig. 7. This figure confirms that Kinect v2 is a very powerful tool which can be applied not only for the versatile management of videogames but also for the display of 3-D images with full parallax, good lateral depth, and for a long range of axial distances.

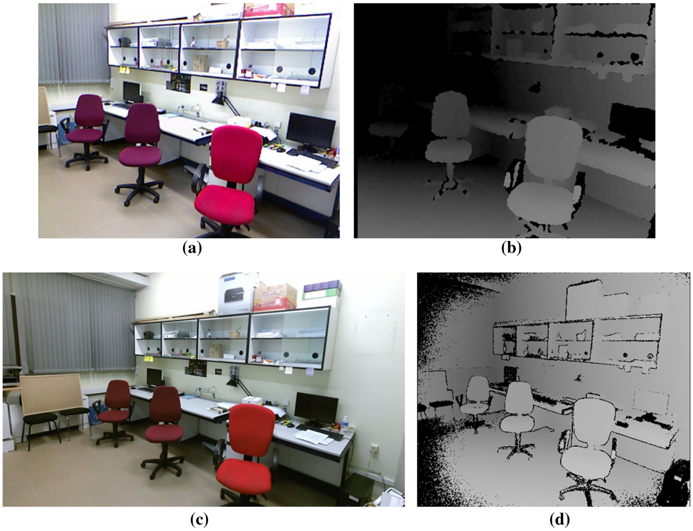


Fig. 3 Captured images from two versions of Kinect: (a, b) Kinect v1 and (c, d) Kinect v2. Both pairs of images are captured from the same standpoint.

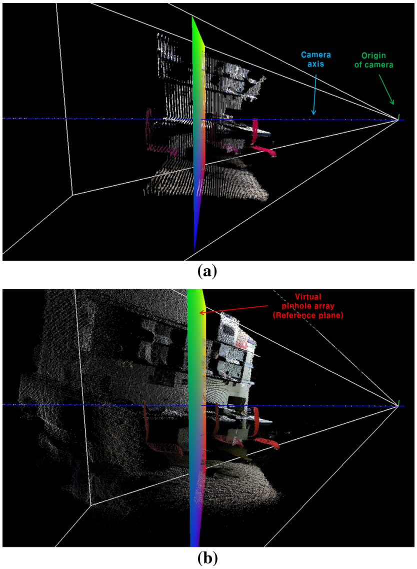


Fig. 4 Display the 3-D point cloud into a virtual 3-D space: (a) from Kinect v1 and (b) from Kinect v2. In both cases, the reference plane is located just behind the second chair.

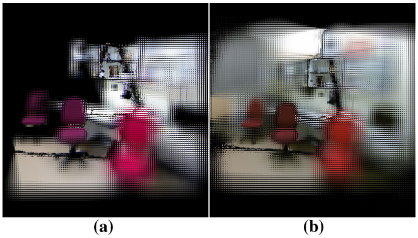


Fig. 5 Collection of microimages generated from modified 3-D point cloud: (a) microimage from Kinect v1 and (b) is from Kinect v2. Both (a) and (b) are generated based on a given reference plane.

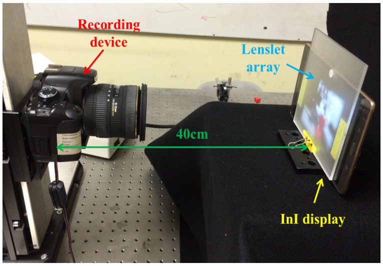


Fig. 6 The overview of our experimental system. We moved the recording device vertically and horizontally to record different perspectives of an integrated image from the proposed InI display system.

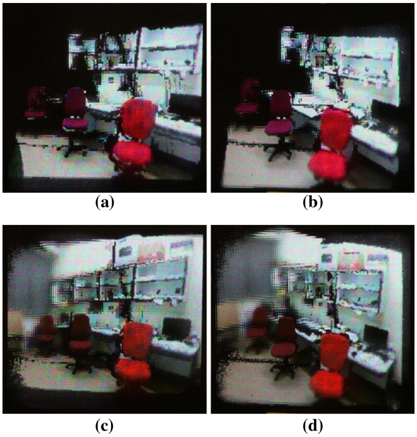


Fig. 7 Images displayed from two different versions of the Kinect: (a, b) Kinect video 1 and (c, d) Kinect video 2. (a, c) Left-bottom view and (b, d) right-top view from proposed InI display (Video 1, mp4, 9.07 MB) [URL: http://dx.doi.org/10.1117/1.OE.56.4.041305.1] and (Video 2 mp4, 10.6 MB) [URL: http://dx.doi.org/10.1117/1.OE.56.4.041305.2].

5 Conclusion

We have reported a comparison of 3-D InI display based on two different versions of Kinect, and demonstrated that Kinect v2 is fully adapted for the task of capturing 3-D optical information for 3-D display. Specifically, we have demonstrated that an InI monitor injected with the information calculated from Kinect v2 data has the ability of displaying 3-D images in color for big scenes. The images have good lateral and depth resolution and also a long range of axial distances. The main drawback of this technique is the existence of black-pixel areas, which result from the capture from a single perspective. In a future work, we will combine the information captured with more than one Kinect v2, in order to obtain a 3-D point cloud that is denser and free of perspective holes.

Acknowledgments

This work was supported by the Ministerio de Economía y Competitividad, Spain (Grant No. DPI 2015-66458-C2-1-R), and by Generalitat Valenciana, Spain (project PROMETEOII/2014/072). S. Hong acknowledges a predoctoral grant from University of Valencia (UV-INV-PREDOC15-265754).

References

1. A. Stem et al., "Three-dimensional image sensing, visualization, and processing using integral imaging," *Proc. IEEE* **94**, 591–607 (2006).
2. J. Jang et al., "Improved viewing resolution of three-dimensional integral imaging by use of non-stationary micro-optics," *Opt. Lett.* **27**, 324–326 (2002).
3. F. Okano et al., "Real-time pickup method for a three-dimensional image based on integral photography," *Appl. Opt.* **36**, 1598–1603 (1997).
4. Y. Kim et al., "Recent researched based on integral imaging display method," *3D Res.* **1**, 17–27 (2010).
5. S. Hong et al., "Viewing angle-improved 3D integral imaging display with eye tracking sensor," *J. Inf. Commun. Converg. Eng.* **12**, 208–214 (2014).
6. S. Hong et al., "Towards 3D television through fusion of Kinect and integral-imaging concepts," *J. Display Technol.* **11**, 894–899 (2015).
7. A. Dorado et al., "Computation and display of 3D movie from a single integral photography," *J. Display Technol.* **12**, 695–700 (2016).
8. G. Park et al., "Multi-viewer tracking integral imaging system and its viewing zone analysis," *Opt. Express* **17**, 17895–17928 (2009).
9. J. Zhang et al., "Integral imaging display for natural scene based on KinectFusion," *Optik* **127**, 791–794 (2016).
10. X. Xiao et al., "Advances in three-dimensional integral imaging: sensing, display, and applications," *Appl. Opt.* **52**, 546–560 (2013).
11. M. Cho et al., "Three-dimensional optical sensing and visualization using integral imaging," *Proc. IEEE* **99**, 556–575 (2011).
12. Lytro, Lytro camera, 2011 https://www.lytro.com.
13. Pelican Imaging, PiCam: Pelican Imaging Camera, 2013 http://www.pelicanimaging.com.
14. Raytrix, Raytrix camera, 2010 http://www.raytrix.de.
15. Canesta, Canesta sensor, 2002 http://en.wikipedia.org/wiki/Canesta.
16. Point Gray, Bumblebee2, 2006 http://www.ptgrey.com/stereo-vision-cameras-systems.
17. Stereolabs, ZED, 2015 http://www.stereolabs.com.
18. Microsoft, "Kinect for windows sensor components and specifications," 2013 http://msdn.microsoft.com/en-us/library/jj131033.aspx.
19. Primesense, Calmine 1.08 & 1.09, 2013 http://en.wikipedia.org/wiki/PrimeSense.
20. Microsoft, "Kinect for Xbox one components and specifications," 2013 http://dev.windows.com/en-us/kinect/hardware.
21. Intel, Realsense, 2014 https://software.intel.com/en-us/realsense.
22. D. Pagliari et al., "Calibration of Kinect for Xbox one and comparison between the two generations of Microsoft sensors," *Sensors* **15**, 27569–27589 (2015).
23. R. Smeenk, "Kinect v1 and Kinect v2 fields of view compared," 2014 http://smeenk.com/kinect-field-of-view-comparison.
24. D. LeMaster et al., "Mid-wave infrared 3D integral imaging at long range," *J. Display Technol.* **9**, 545–551 (2013).
25. G. R. Bradski et al., *Learning OpenCV: Computer Vision with the OpenCV Library*, O' Reilly Media Press, California (2008).
26. R. Sukthankar et al., "Smarter presentations: exploiting homography in camera-projector systems," in *Proc. of Int. Conf. on Computer Vision*, pp. 247–253 (2001).
27. Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *The Proc. IEEE 7th Int. Conf. on Computer Vision*, pp. 666–673 (1999).
28. Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 1330–1334 (2000).
29. OpenCV, OpenCV: camera calibration, 2000, http://docs.opencv.org/doc/tutorials/calib3d/camera_calibration/camera_calibration.html.
30. H. Shin et al., "Hybrid stereoscopic camera system," *J. Broadcast Eng.* **16**, 602–613 (2011).
31. M. Martinez-Corral et al., "Formation of real, orthoscopic integral images by smart pixel mapping," *Opt. Express* **13**, 9175–9180 (2005).

Seokmin Hong received his BEng and MSc degrees in digital and visual contents from Dongseo University, Busan, Korea, in 2012 and 2014, respectively. In 2012, Dongseo University honored him with the BEng Extraordinary Award. Since 2015, he has been working with the 3-D Imaging and Display Laboratory, Optics Department, University of Valencia, Spain. His research interests are image processing, computer vision, and applied computer science.

Genaro Saavedra received his BSc and PhD degrees in physics from the Universitat de València, Spain, in 1990 and 1996, respectively. His PhD work was honored with the PhD Extraordinary Award. He is currently a full professor of optics at this university, where he is coleader of the 3-D Imaging and Display Laboratory. His current research interests are optical diffraction, integral imaging, 3-D high-resolution optical microscopy, and phase-space representation of scalar optical fields.

Manuel Martinez-Corral received his MSc and PhD degrees in physics from the University of Valencia in 1988 and 1993, respectively. In 1993, the University of Valencia honored him with the PhD Extraordinary Award. He is currently a full professor of optics at the University of Valencia, where he is coleader of the 3-D Imaging and Display Laboratory. His research interest includes resolution procedures in 3-D scanning microscopy, and 3-D imaging and display technologies.

Paper IV

Full-parallax 3D display from stereo-hybrid 3D camera system

Seokmin Hong, Amir Ansari, Genaro Saavedra, and Manuel Martínez-Corral

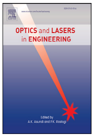
in *Optics and Lasers in Engineering* 103, 46–54 (2018).

Optics and Lasers in Engineering 103 (2018) 46–54

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng



Full-parallax 3D display from stereo-hybrid 3D camera system

Seokmin Hong*, Amir Ansari, Genaro Saavedra, Manuel Martinez-Corral

University of Valencia, 3D Imaging and Display Laboratory, Department of Optics, Burjassot, Valencia, E-46100, Spain

ARTICLE INFO

Keywords:
3D display
Integral imaging
3D data registration
Color transfer
Point cloud
Stereo-hybrid 3D camera

ABSTRACT

In this paper, we propose an innovative approach for the production of the microimages ready to display onto an integral-imaging monitor. Our main contribution is using a stereo-hybrid 3D camera system, which is used for picking up a 3D data pair and composing a denser point cloud. However, there is an intrinsic difficulty in the fact that hybrid sensors have dissimilarities and therefore should be equalized. Handled data facilitate to generating an integral image after projecting computationally the information through a virtual pinhole array. We illustrate this procedure with some imaging experiments that provide microimages with enhanced quality. After projection of such microimages onto the integral-imaging monitor, 3D images are produced with great parallax and viewing angle.

1. Introduction

During the last century, the three-dimensional (3-D) imaging systems have been issued in order to record and display 3-D scenes. Among them, integral-imaging (InI) has been considered as one of the prospective technologies in order to reflect real 3-D scenes into a multi-visual display system. This concept was proposed by G. Lippmann in 1908. He presented the possibility of capturing the 3-D information and reconstructing the 3-D scene by using spherical diopter arrays [1–3]. Depending on its manipulation, InI is classified by two stages: pickup and display. Nowadays, the pickup procedure is performed by placing a tiny lens array in front of a two-dimensional (2-D) imaging sensor and producing the collection of microimages. A noteworthy feature is that every microimage contains different perspective information. This is because all of the light rays reflected (or diffused) by an object are transmitted by all the lenses, which distribute the light on different pixels of the microimages depending on the incidence angle. Hereafter, the whole array of microimages is referred to as the integral image. Concerning the display stage, when the integral image is projected onto an InI display system, observers can see the 3-D floating color scene, which has full-parallax and quasi-continuous perspective view [4–7]. Many researchers and companies have applied the InI technique in many different fields [8–18].

In the meantime, various depth-sensing techniques were launched in order to record 3-D scenes [19–25]. Among all, one of highlighted techniques is stereovision, which takes advantage of the disparity information from two aligned cameras which has been the representative of the depth-image sensing for a long period [19–20]. Incidentally, in the past decades, the use of technologies related to infrared (IR) light sensors has become spotlighted [21–25]. Especially the Kinect device from Microsoft takes profit from IR lighting technology in the case of depth acquisition. By this time, two different versions of the Kinect are released. The main commercial specifications of Kinect v1 (Kv1) and v2 (Kv2) are described in Table 1. These devices allow to acquire RGB images, IR images and also depth information in real-time with a high frame rate. As well known, both devices have many different features for obtaining a dense depth map. Kv1 uses a structured IR light pattern emitter and IR camera to calculate the depth distance through the captured pattern [21–23]. In contrast, Kv2 utilizes time-of-flight (ToF) technology, which exploits emitting IR beams with high frequency. Having the reflected IR light from most 3-D surfaces, the sensor evaluates the depth distance by measuring the IR flash's returning duration [24–25].

In a previous work, we proposed the use of RGB image and depth information obtained by a single 3-D camera to generate an integral image and project it onto an InI display system [13–15]. However, this innovative approach still contains several issues that must be improved. Among them, the main drawbacks are domination of the depth information by the noise caused by the limitation of IR light sensing technique; the low density of depth map, which is restricted by the sensor's specification; and the depth-hole problem, which occurs because of the reflections and/or occlusions. Mono-perspective devices can see only the frontal part of scenes, so that occluded area's information is lost in the scene.

* Corresponding author.
E-mail address: seokmin.hong@uv.es (S. Hong).

https://doi.org/10.1016/j.optlaseng.2017.11.010
Received 21 June 2017; Received in revised form 19 October 2017; Accepted 13 November 2017
Available online 23 November 2017

Table 1
Comparison between Kv1 and Kv2 specifications.

List	Kinect v1	Kinect v2
Released (year)	2010	2014
RGB camera (pixel)	640 × 480 (Max: 1280 × 960)	1920 × 1080
Frames per second in RGB camera	30 (Max: 12)	30 (low-light condition: 15)
IR camera (pixel)	640 × 480	512 × 424
Frames per second in IR camera	30	30
Depth acquisition method	Structured IR light pattern	Time of Flight
Suitable depth range (mm)	800–4000	500–4500
IR camera's Horizontal FOV (°)	57	70
IR camera's Vertical FOV (°)	43	60

Table 2
Calibrated camera parameters to calculate the scale factor between target sensors.

Sensor	Coordinate (u: width, v: height)	Resolution (# of pixels)	Sensor size (mm)	Pixels per mm	Focal length (mm)
Kinect v1 (RGB camera)	u	640	3.58	178.771	3.099
	v	480	2.87	167.247	
Kinect v2 (IR camera)	u	512	5.12	100	3.657
	v	424	4.24	100	

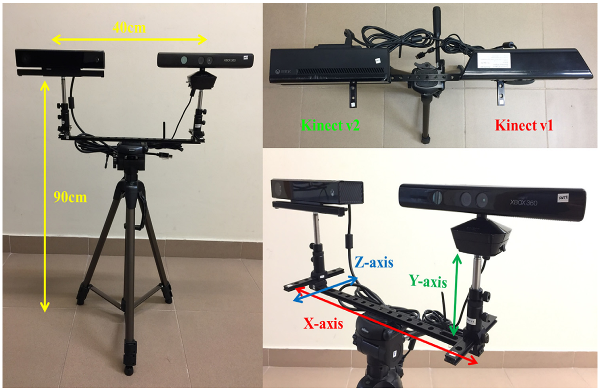


Fig. 1. Proposed stereo-hybrid 3-D camera system.

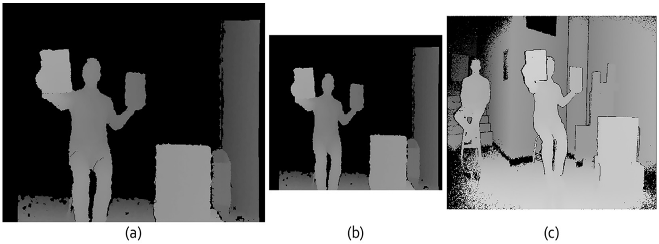
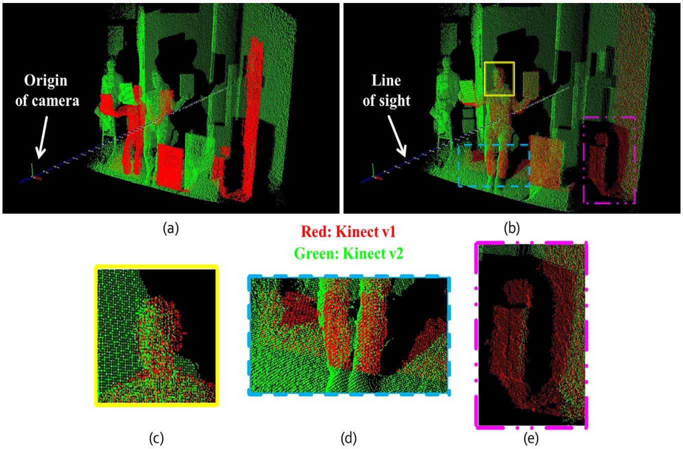


Fig. 2. Captured depth maps from our experimental camera system: (a) from Kv1, (b) rescaled image; and (c) from Kv2.

In order to solve these limitations, we propose the use of stereo-hybrid 3-D camera system. Fig. 1 shows our camera setting. In comparison to monocular system, stereo-vision system can generally extend the field of view (FOV) against with. In order words, our approach can expand the visual space and obtain the occluded information taking advantage of binocular system. Thus, depth-hole area is filled in by complementing each other. Another important advantage of the proposed method is yielding denser point cloud. This procedure is described in Section 2. With this improved 3-D data, the microimages are generated with higher quality. The microimages generation process is described



(a) (b)

(c) (d) (e)

Fig. 3. 3-D point clouds in the virtual 3-D space: (a) before registration process; (b) after calculation result; in (c–e) we magnified some specific parts of the scene. In the figures, red color point is rescaled point clouds from Kv1, and green color point is from Kv2 respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in Section 3. Finally, in Sections 4 and 5 we provide our experimental results and conduct the conclusions respectively.

2. Stereo-Hybrid point clouds manipulation

In order to implement the stereo system, it is convenient either the use of two 3-D cameras of the same model, or the use of two different 3-D cameras with complementary features. In our approach we have decided to make use of Kinect technology. To the best of our knowledge, high frame rate synchronization of two Kv2 nor two Kv1 is never addressed so far. Hence at this stage of our research we have decided to tackle the implementation of a stereo-hybrid technique by taking profit of complementary features of Kv1 and Kv2. Note that even in the case of 2-D cameras, it has been very unusual to compose hybrid camera systems [26–28]. This is an evident motivation why we want to use hybrid 3-D cameras into our research, since its outcomes can be very useful for a potential manipulation of various types of cameras in further research. In the Section 2.1, the correction of the different scale information between sensors will be explained. In sequence, the arrangement and registration of the individual 3-D point cloud information will be shown. In Section 2.2, the correction of the color dissimilarity of sensors will be presented.

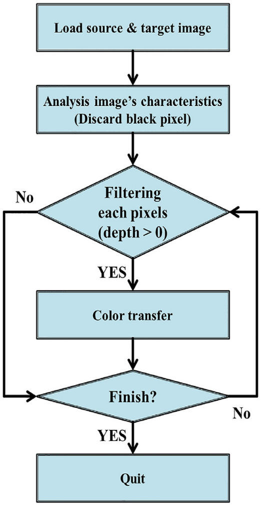

Load source & target image
Analysis image's characteristics (Discard black pixel)
Filtering each pixels (depth > 0)
No
YES
Color transfer
Finish?
No
YES
Quit

Fig. 4. Flow chart of the proposed color transferring strategy. The loop is applied voxel by voxel.

2.1. Hybrid point clouds registration

In our previous paper [15], we mentioned about the difference between Kv1 and Kv2. Above all, each Kinect devices has two camera sensors (RGB and IR) by its own, and the four sensors have different FOV and image resolution. It means that all of them have their own scale factors, which need to be corrected. In [26], authors proposed how to correct the scale information in hybrid stereoscopic 2-D camera systems. The algorithm manages the images captured by two different sensors, the input image and the target image, and aims to obtain a rescaled input image. Eqs. (1) and (2) show how to derive scale factors: $i_{u,\,v}$ refers to number of pixels in the input image, while $j_{u,\,v}$ is the number of pixels in the corrected input image. Besides, f, f' are input and target focal lengths (mm), whereas $p_{u,\,v}$, $p'_{u,v}$ are input and target pixels/mm.

$$\begin{cases} j_u = \dfrac{i_u f' p'_u}{f\, p_u} = \lambda_u i_u \\ j_v = \dfrac{i_v f' p'_v}{f\, p_v} = \lambda_v i_v \end{cases} \tag{1}$$

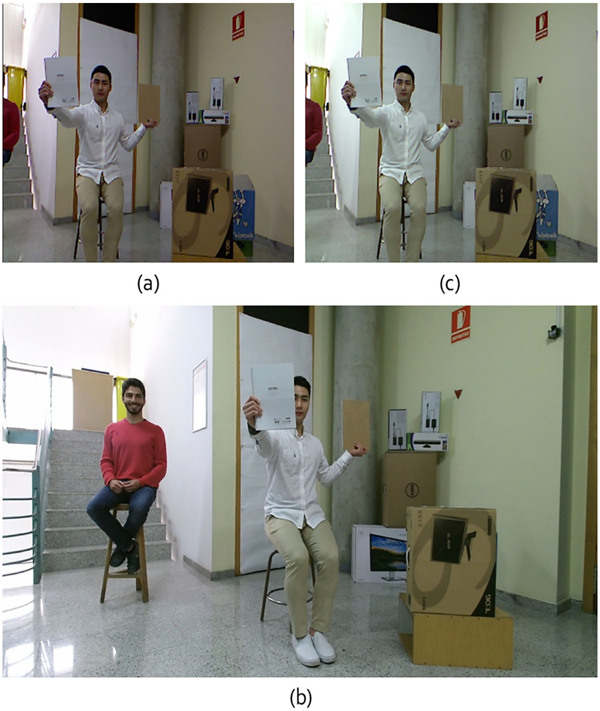


Fig. 5. Processed result by using referenced color transfer algorithm (a) rescaled RGB image captured by Kv1 (b) the captured RGB image from Kv2 (c) color transferred result from (b) to (a).

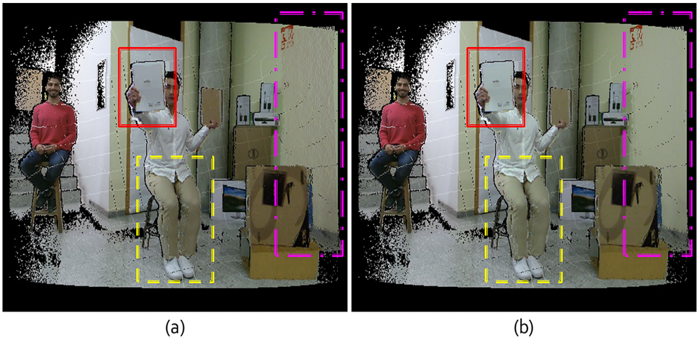


Fig. 6. Orthographic projection of registered 3-D point cloud: (a) before color transferring process (b) is after color transfer. Note that (b) shows more natural textured scene than (a).At the scene, the black pixels have no information because they are out of the depth-range capacity of IR sensing.

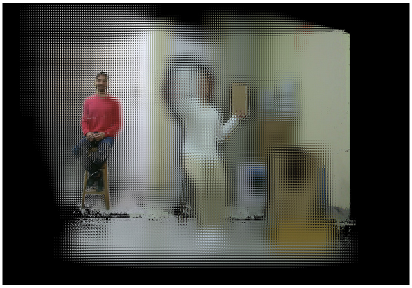


Fig. 7. Collection of microimages generated from modified 3-D point cloud. In this case, the reference plane is placed in 2450 mm distance from the origin of Kv2.

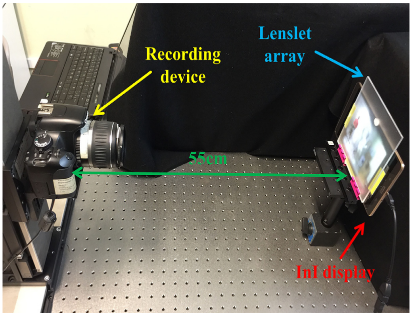


Fig. 8. Overview of experimental system.

Where we have defined parameter

$$\lambda_{u,v} = \frac{f' p'_{u,v}}{f\, p_{u,v}} \tag{2}$$

as the scale factor between target and input sensors.

In our approach, input image is Kv1's RGB image and target image is Kv2's IR image. There are several reasons why we decided to transpose from Kv1's RGB camera to Kv2's IR camera. First, mapping from IR image to RGB image in Kv1 is feasible because of many solutions are already released. Second, Kv2 depth information is denser and with larger FOV than Kv1. Finally, the third reason is that in the Kv2, the resolution in RGB camera is much bigger than that in IR camera. If we would map from IR image to RGB image, in Kv2, IR image needs, not only to up-scaling, but also interpolate the pixel gaps in rescaled image.

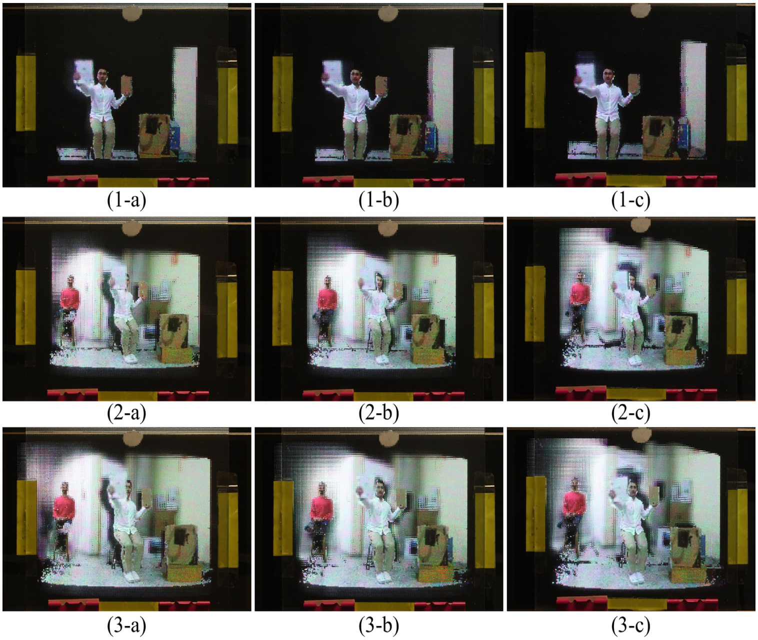


Fig. 9. Comparison result between displayed integral image: (1-a, b, c) Kv1, (2-a, b, c) Kv2, (3-a, b, c) the proposed result. (a, b, c) shows different perspective position where (a) is left-bottom, (b) is right-bottom, and (c) is right-top from the InI monitor. All images are excerpted from recorded video: media 2, 3, and 4.

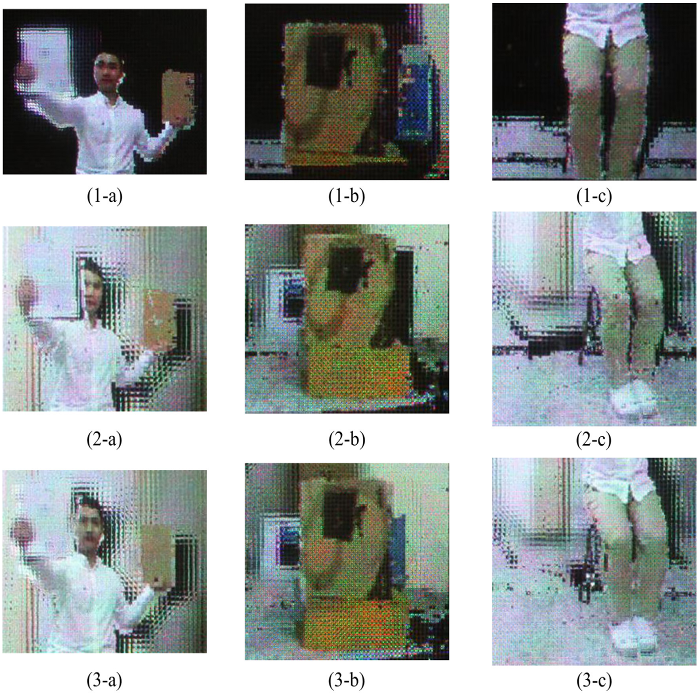


Fig. 10. More detail comparison of displayed integral image. This figure shows the advantage of our approach. We filled in several depth-hole areas and derived to smoother texture at the scene. Above all, some occluded area is recovered precisely.

To calculate the scale factor, the well-treated data from [29] is followed, where both Kinect parameters were calibrated accurately (see Table 2). Then the scale factors are, $\lambda_u = 0.660$, $\lambda_v = 0.706$; and rescaled input resolutions are $j_u \cong 422.46, j_v \cong 338.68$. Fig. 2 shows captured depth map from Kv1 and Kv2's IR sensors and rescaled result. See the figures for further details.

From now on, rescaled image resolution of Kv1's RGB and Kv2's IR camera are adapted to the same scale information. Afterward, the captured RGB and Depth information from each device can compose point cloud and dispose into a virtual 3-D space. However, both cloud data are still mutually shifted and not arranged properly (See Fig. 3(a)). In order to make registration between two point cloud sets, Iterative-Closest-Point (ICP) algorithm is utilized. ICP algorithm calculates the movement between two sets of point clouds in order to minimize their distance. ICP is often used to reconstruct 2-D or 3-D data captured from different positions. The output of ICP algorithm is rigid (or rigid body) transformation matrix, which includes translation and rotation [30–32]. Fig. 3 shows the point cloud before and after registration result. The red and green colors represent the point cloud obtained by Kv1 and Kv2 respectfully. As it can be seen, the Kv1's data are well-aligned into Kv2 and covered in some occluded area. Especially, Fig. 3(c–e) indicates more detail of the registration result clearly.

2.2. *Color transfer between color images*

Even when the two point clouds are registered properly, the RGB images of the Kv1 and Kv2 still have color dissimilarities. To overcome this drawback, the color transfer method proposed by Reinhard et al. [33] is followed, but it is adapted to 3-D images. Our approach is described in the flowchart of Fig. 4. In the second step, after loading the input and target point clouds, the black voxels having no color information are discarded. Then, the voxels without depth information were filtered out. The reason for such discarding is that those meaningless voxels would transfer wrong color characteristics. As result of applying the algorithm to all the voxels, the Kv2's RGB color values are transferred onto the characteristics of the Kv1's RGB image. Figs. 5 and 6 show the color-transfer result clearly. In Fig. 5(a) we show the input RGB image (obtained with Kv1), in Fig. 5(b) the target RGB image (Kv2) and finally in Fig. 5(c) the modified input image after the color transference.

Fig. 6 shows orthographic projection of RGB information of registered 3-D point clouds. Fig. 6(a) shows the point cloud before the color transfer, while Fig. 6(b) shows the same point cloud after the transfer. In Fig. 6, the areas of the scene where significant improvement is obtained due to the color transfer have been marked. In order to illustrate and demonstrate our proposal, the video Media 1 is composed with this sequence: point clouds of Kv1, Kv2, without registration result, and registration with color transfer result respectively.

3. Microimages generation from point clouds

In order to generate the microimages for their projection onto the InI display system, our previous approach [14] is followed. Then in our algorithm we placed a virtual pinhole array (VPA) at a certain distance from the 3-D point cloud. Indeed, VPA's location reflects the correlation between real scene and displayed scene. In particular, this position determines the front and rear volumes at the displayed 3-D scene. Accordingly, we entitle this position as a reference plane. Then, the voxels of each point cloud are projected through the VPA, so that the integral image is composed, as in [34]. In this back-projection scheme, each microimage records the angular information. In fact, the calculation of the microimages needs to account for the parameters of the InI display system; i.e. the number of microlenses, their pitch, gap, and number of pixels behind any microlens. Fig. 7 shows calculated microimages, which are ready for projection through the InI display.

4. Experimental results of displayed three-dimensional image

In our experiment, the InI monitor is composed of a Samsung SM-T700 (359 pixels/inch) tablet, and a MLA consisting of 181 × 113 lenslets of focal length $f = 3.3$ mm and pitch $p = 1.0$ mm (Model 630 from Fresnel Technology). Each microimage is composed of 15 × 15 pixels, the gap between the microlenses and the display is fixed to $g = 49.5$ pixels, and thus, the full size of the integral image is 2715 × 1695 pixels (14.13 pixels/mm). After mounting and aligning the MLA in front of the tablet, the 3-D scene is displayed with full-parallax.

To demonstrate the proposed approach, the setup is implemented as shown in Fig. 8. The InI monitor displays and integrates the microimages towards the observer's eyes. Originally, our target is binocular observers, who can see the 3-D nature of displayed scene, that is, they can perceive several parts of the displayed scene in front of the monitor and some others behind. Unfortunately, this full-parallax effect cannot be directly demonstrated in a manuscript or even in a monocular video. In order to demonstrate this effect we proceeded as follows. First the observer is replaced by a monocular digital camera. Then a collection of pictures is obtained after displacing horizontally and vertically the camera along a region of 70 × 70 mm. With these pictures, a video is composed in which the InI monitor was observed from different perspectives. Media 2 and 3 shows Kv1 and Kv2, and Media 4 shows the final modified result. All of the recorded videos are composed of different horizontal and vertical perspective views. The Figs. 9 and 10 show this experimental result more clearly. Modified point clouds are filled in some depth-hole areas and as a result, it induces denser and smoother texture of the scene. The most impressive feature is that some occluded areas are recovered by registration process. Especially, the human model's head and blue basket behind of brown box are recovered properly.

5. Conclusion

To the best of our knowledge, this is the first time to utilize a stereo-hybrid 3-D camera system to capture the light field. Specifically, in order to overcome the limitations of a mono perspective view, the usage of stereo-hybrid system consisting of two Kinect devices is proposed. But we had to tackle the challenge of fusing two different 3-D point clouds with strong dissimilarities: different lateral and axial resolution, different spectral sensitivities of RGB sensors, and even different luminance of the 3-D scene when seen from different perspectives. To cope with these mismatches, some well-known algorithms fitting to our specific situation have been adapted. To demonstrate our approach, a 3-D scene is captured with the stereo-Kinect device and the 3-D point clouds are modified according to our strategy of correcting the dissimilarities. Finally, the improvements in the displayed images have been demonstrated by calculating the microimages and projecting them onto an InI monitor, which provides the observers with full-parallax 3-D images. Since we filled in several depth-hole areas at the integral image and derived to smoother texture at the scene, the experiment confirms that the quality of 3-D data is improved noticeably. Above all, some occluded field is recovered precisely and thus, this output proves the benefit of our manipulation. In a future work, we will apply this technique for different and/or newer types of 3-D cameras: Light-field camera [8–11], and stereo-vision camera [19–20]. In addition, we will enhance the accuracy of 3-D data registration and color equalization result. Finally, we would like to point out that a different experimental concept which is manipulated by LeMaster et al. [35], where they use an array of mid-wave infrared cameras to obtain depth reconstructions for long distances is also complementary as our experiment.

Acknowledgement

This work was supported by the Ministerio de Economía y Competitividad, Spain (Grant no. DPI 2015-66458-C2-1R), and by Generalitat Valenciana, Spain (project PROMETEOII/2014/072). S. Hong acknowledges a predoctoral grant from University of Valencia (UV-INVPREDOC15-265754). And A. Ansari acknowledges a predoctoral contract from EU H2020 program under MSCA grant 676401.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.optlaseng.2017.11.010.

References

[1] Lippmann G. Epreuves réversibles photographies integrals. Comptes Rendus de l'Académie des Sciences 1908;146:446–51.
[2] Lippmann G. Epreuves reversibles donnant la sensation du relief. J Phys Théorique et Appl 1908;7:821–5.
[3] Lippmann G. L'étalon international de radium. Radium (Paris) 1912;9:169–70.
[4] Stem A, Javidi B. Three-dimensional image sensing, visualization, and processing using integral imaging. Proc IEEE 2006;94:591–607.
[5] Okano F, Hoshino H, Arai J, Yuyama I. Real-time pickup method for a three-dimensional image based on integral photography. Appl Opt 1997;36:1598–603.
[6] Kim Y, Hong K, Lee B. Recent researched based on integral imaging display method. 3D Res 2010;1:17–27.
[7] Park S, Yeom J, Jeong Y, Chen N, Hong J, Lee B. Recent issues on integral imaging and its applications. J Inf Disp 2014;15:37–46.
[8] Lytro camera, 2011. https://www.lytro.com.
[9] PiCam: Pelican Imaging Camera, 2013. http://www.pelicanimaging.com.
[10] Raytrix camera, 2010. http://www.raytrix.de.
[11] Canesta sensor, 2002. http://en.wikipedia.org/wiki/Canesta.
[12] Ng R, Levoy M, Bredif M, Duval G, Horowitz M, Hanrahan P. Light field photography with a hand-held plenoptic camera. Light field photography with a hand-held plenoptic camera, 2; 2005. Tech. Rep. CSTR. 2.
[13] Hong S, Shin D, Lee J, Lee B. Viewing angle-improved 3D integral imaging display with eye tracking sensor. J Inf Commun Converg Eng 2014;12:208–14.
[14] Hong S, Shin D, Lee B, Dorado A, Saavedra G, Martinez-Corral M. Towards 3D television through fusion of kinect and integral-imaging concepts. J Disp Technol 2015;11:894–9.
[15] Hong S, Saavedra G, Martinez-Corral M. Full parallax three-dimensional display from Kinect v1 and v2. Opt Eng 2016;56:041305.
[16] Dorado A, Saavedra G, Sola-Pikabea J, Martinez-Corral M. Integral imaging monitors with an enlarged viewing angle. J lnf Commun Converg Eng 2015;13:132–8.
[17] Dorado A, Martinez-Corral M, Saavedra G, Hong S. Computation and display of 3D movie from a single integral photography. J Disp Technol 2016;12:695–700.
[18] Han Y, Lee M, Lee B. Air-touch interaction system for integral imaging 3D display. In: First Int. Workshop on Pattern Recognit.; 2016. p. 10011.
[19] Bumblebee2, 2006. http://www.ptgrey.com/stereo-vision-cameras-systems.
[20] ZED, 2015. http://www.stereolabs.com.
[21] Kinect for windows sensor components and specifications; 2013. http://msdn.microsoft.com/en-us/library/jj131033.aspx.
[22] Primesense Calmine 1.08 & 1.09, 2013. http://en.wikipedia.org/wiki/PrimeSense.
[23] ASUS Xtion, 2011. http://www.asus.com/3D-Sensor/Xtion_PRO.
[24] Kinect for Xbox one components and specifications; 2013. http://dev.windows.com/en-us/kinect/hardware.
[25] Intel Realsense, 2014. https://software.intel.com/en-us/realsense.
[26] Shin H, Kim S, Sohn K. Hybrid stereoscopic camera system. J Broadcast Eng 2011;16:602–13.
[27] Garcia F, Aouada D, Mirbach B, Ottersten B. Real-time distance-dependent mapping for a hybrid ToF multi-camera rig. IEEE J Sel Top Signal Process 2012;6:425–36.
[28] Frick A, Koch R. LDV generation from multi-view hybrid image and depth video. In: 3D-TV system with depth-image-based-rendering. New York: Springer; 2012. p. 191–220.

[29] Pagliari D, Pinto L. Calibration of kinect for Xbox one and comparison between the two generations of microsoft sensors. Sensors 2015;15:27569–89.

[30] Besl PJ, Mckay ND. A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell 1992;14:239–56.

[31] Zhang Z. Iterative point matching for registration of free-form curves and surfaces. Int J Comput Vis 1994;13:119–52.

[32] Rusinkiewicz S, Levoy M. Efficient variants of the ICP algorithm. In: Proceedings Third Int. Conf. on 3-D Digit. Imaging and Mode; 2001. p. 145–52.

[33] Reinhard E, Ashikhmin M, Gooch B, Shirley P. Color transfer between images. IEEE Comput Graph Appl 2001;21:34–41.

[34] Martinez-Corral M, Javidi B, Martinez-Cuenca R, Saavedra G. Formation of real, orthoscopic integral images by smart pixel mapping. Opt Express 2005;13:9175–80.

[35] LeMaster D, Karch B, Javidi B. Mid-wave infrared 3D integral imaging at long range. J Disp Technol 2013;9:545–51.

Seokmin Hong received the B.Eng. and M.Sc. degrees in digital and visual contents from Dongseo University, Busan, South Korea, in 2012 and 2014, respectively. In 2012, Dongseo University honored him with the B.Eng. Extraordinary Award. Since 2015, he has been working with the 3D Imaging and Display Laboratory, Optics Department, University of Valencia, Spain. His research interests are image processing, computer vision, and applied computer science.

Amir Ansari received his master degree in Electrical Engineering-Communication systems from Shiraz University of Technology, Iran in 2014. As a master student, he studied signal/image processing, digital watermarking, image and video coding and his research activity was mainly deviated to hyperscectral image fusion. He also has experience of designing GPS-based remote positioning and telemetry systems. On September 2016, he qualified for Marie Sklodowska Curie scholarship and joined 3DID where he is currently doing his PhD. His research interests include (but are not limited to) image processing/watermarking/fusion, hardware design and navigation systems.

Genaro Saavedra received the B.Sc. and Ph.D. degrees in physics from Universitat de València, Spain, in 1990 and 1996, respectively. His Ph. D. work was honored with the Ph.D. Extraordinary Award. He is currently Full Professor with Universitat de València, Spain. Since 1999, he has been working with the “3D Display and Imaging Laboratory”, at the Optics Department. His current research interests are optical diffraction, integral imaging, 3D high-resolution optical microscopy and phase-space representation of scalar optical fields. He has published on these topics about 50 technical articles in major journals and 3 chapters in scientific books. He has published over 50 conference proceedings, including 10 invited presentations.

Manuel Martínez-Corral was born in Spain in 1962. He received the M.Sc. and Ph.D. degrees in physics from the University of Valencia, Spain, in 1988 and 1993, respectively. In 1993, the University of Valencia honored him with the Ph.D. Extraordinary Award. He is currently Full Professor of Optics at the University of Valencia, where he is co-leader of the “3D Imaging and Display Laboratory”. His research interest includes resolution procedures in 3D scanning microscopy, and 3D imaging and display technologies. He has supervised on these topics seven Ph. D. theses, two of them honored with the Ph.D. Extraordinary Award. He has published over eighty technical articles in major journals, and pronounced over thirty invited and five keynote presentations in international meetings. He has been member of the Scientific Committee in over twenty international meetings. In 2010, Dr. Martinez-Corral was named Fellow of the SPIE. He is co-chair of the Three-Dimensional Imaging, Visualization, and Display Conference within the SPIE meeting in Defense, Security, and Sensing (Baltimore). He was Topical Editor of the IEEE/OSA JOURNAL OF DISPLAY TECHNOLOGY.

Paper V

New Method of Microimages Generation for 3D Display

Nicolò Incardona, Seokmin Hong, Manuel Martínez-Corral, and Genaro Saavedra

in *Sensors* 18, 2805 (2018).

sensors

MDPI

Article

New Method of Microimages Generation for 3D Display

Nicolò Incardona *, Seokmin Hong, Manuel Martínez-Corral and Genaro Saavedra

Department of Optics, University of Valencia, 46100 Burjassot, Valencia, Spain; seokmin.hong@uv.es (S.H.); manuel.martinez@uv.es (M.M.-C.); genaro.saavedra@uv.es (G.S.)
* Correspondence: nicolo.incardona@uv.es

Received: 4 August 2018; Accepted: 23 August 2018; Published: 25 August 2018

Abstract: In this paper, we propose a new method for the generation of microimages, which processes real 3D scenes captured with any method that permits the extraction of its depth information. The depth map of the scene, together with its color information, is used to create a point cloud. A set of elemental images of this point cloud is captured synthetically and from it the microimages are computed. The main feature of this method is that the reference plane of displayed images can be set at will, while the empty pixels are avoided. Another advantage of the method is that the center point of displayed images and also their scale and field of view can be set. To show the final results, a 3D InI display prototype is implemented through a tablet and a microlens array. We demonstrate that this new technique overcomes the drawbacks of previous similar ones and provides more flexibility setting the characteristics of the final image.

Keywords: integral imaging; plenoptic imaging; 3D display; CGII

1. Introduction

3D TV implementation is a fascinating challenge for the researchers of many different scientific communities. The first generation of 3D TV is no longer produced: one of the main drawbacks of these devices was the need of glasses in order to see the 3D content. For this reason, a new type of glasses-free device is being investigated, the so-called autostereoscopic displays. Among these, multi-view displays allow 3D visualization for multiple viewers with stereo and movement parallax, and overcome the accommodation-convergence conflict [1,2]. This kind of displays are named Integral Photography (IP) or Integral Imaging (InI) displays, because their operating principle is based on the Integral Photography technique. It was proposed one century ago by Gabriel Lippmann to register the 3D information of a scene [3]. His idea was to replace the objective lens of the photographic camera with a microlens array (MLA), and to place a photographic film at the focal plane of the lenses. Doing so, different perspectives of the scene are captured. The part of the photographic film (nowadays, the portion of the pixelated sensor) behind each microlens corresponds to a different perspective. These perspective views are called Elemental Images (EIs) and the set of EIs is the Integral Image (InI). The aim of Lippmann was to project the images captured with IP through a MLA similar to the one used in the capturing stage. Doing so, the light emitted by the EIs of the photographic film is integrated in front of the MLA, producing a 3D reconstruction of the original captured scene (Figure 1).

The great progress in optoelectronic technologies renewed the interest in this technique. Commercial cameras based on IP, known as plenoptic cameras, are already available on the market [4]. These cameras give access to a great number of applications such as the extraction of the depth map, or digital refocusing of the picture. Moreover, the continuous advances in displays (4 K and even 8 K displays are already available on the market) and in MLA manufacturing are a great boost for research in 3D InI displays.

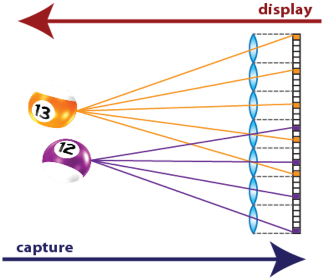


Figure 1. A 3D scene captured and displayed with IP technique. Note that, in the capturing stage, the observer is in the right side (behind the camera's sensor), so he sees the violet ball closer. In the display stage, the observer is in the left side (in front of the MLA of the InI display), so he sees the orange ball closer: the scene is reconstructed with reversed depth.

It is important to remark that the EIs captured through IP technique are not directly projectable in the 3D InI monitor. Firstly, because the MLA used in the capturing process is not the same of that used in the display. Moreover, if EIs are directly projected, the reconstructed scene will be a pseudoscopic version of the original one, which means that the scene is reconstructed with reversed depth, as shown in Figure 1. To solve these problems, some computations have to be made, to convert the EIs into the so-called microimages. The conversion is made by means of a simple transposition, that is, through a pixel resampling [5].

The proposed method generates and processes the EIs to convert them to microimages. It overcomes the drawbacks of previous ones and is applicable to real scenes. The operating principle is very simple: the 3D scene is firstly captured and converted to a virtual point cloud. Any method that permits the extraction of a 2D depth map of the scene can be used for this task. Then, the EIs are synthetically captured from this point cloud through a virtual cameras array. Finally, the EIs are processed and converted to microimages projectable in the 3D InI display. Since the EIs are synthetically captured, we have much more freedom in adjusting the parameters of the cameras array. In this way, it is possible to change the characteristics of the final image and the way the scene is reconstructed by the 3D InI display. Above all, it is possible to adapt the algorithm to any InI display, without having to repeat the capture of the scene. A geometrical model of the system is exploited, which permits directly setting with high precision the reference plane's position and the field of view of the image. The reference plane's position is fundamental because it sets the portion of the 3D scene that will be reconstructed inside and outside the InI monitor, changing the depth sensation of the scene. The part of the scene that is behind the reference plane is reconstructed by the MLA as a virtual image inside the screen, while the part that is beyond the reference plane is reconstructed as a real image, floating in front of the screen.

This paper is organized as follows. In Section 2, a quick review on the previous work on techniques for microimages generation is done. In Section 3, our new technique is described, and the geometrical model of the system is explained. In Section 4, the experimental results obtained are presented. Finally, in Section 5, the achievements of the presented work are summarized.

2. Previous Work

As stated before, the EIs captured with Lippmann's technique are not directly projectable in the InI monitor. Many methods to generate microimages for InI displays have already been reported. Kwon et al. [6], Jiao et al. [7] and Li et al. [8] used different techniques to generate synthetically the

EIs, and then process them to obtain the microimages. Instead, Chen et al. [9] directly captured the microimages putting a virtual pinhole array (VPA) into the 3D model. However, all these methods only process computer-generated 3D models, not real world 3D scenes.

Among the techniques that process real 3D scenes, Navarro et al. [10] and Martínez-Corral et al. [11] proposed the so-called Smart Pseudoscopic-to-Orthoscopic Conversion (SPOC). A collection of EIs is optically captured. Then, through a smart pixel mapping, the microimages are obtained, with the possibility to change the reference plane's position. This method has some limitations. First, if the synthetic aperture method is used to capture the EIs (EIs captured with a single conventional camera mechanically displaced), the capturing stage is very time-consuming. Moreover, the reference plane's position can be set only at determined planes, because just one parameter can be used for this task.

Hong et al. (2015) [12] and Hong et al. (2018) [13] used Kinect cameras to capture the spatial and depth information of the scene. Then, these data are merged into a point cloud. A VPA is used to directly obtain the microimages adjusted to the InI monitor. With this method, the capturing stage is reduced to a single snapshot of the scene. Besides, the amount of data to process is greatly reduced: an RGB image and a depth map are sufficient. However, another issue appears. The VPA used to generate the microimages is set near or directly inside the scene. The point cloud has a finite number of elements: the ones that are close to the VPA have a very big angle with respect to the pinholes, so they do not map onto any pixel. For this reason, this part of information of the scene is lost and large areas with black pixels, that is pixels with no information, appear in the final image.

Piao et al. [14] and Cho and Shin [15] used off-axially distributed image sensing (ODIS) and axially distributed image sensing (ADS), respectively, to extract the color and depth information of the 3D scene. Again, this information is used to compute synthetically the microimages through a VPA.

The proposed technique overcomes black pixels' drawback and it offers much more flexibility than the mentioned ones.

3. Proposed Technique

The basic idea is to capture the real 3D scene and convert it to a virtual point cloud in order to process it synthetically. A set of EIs of the virtual 3D scene (the point cloud) is captured through a simulated cameras array, and processed as in [11] to obtain the microimages. The EIs are synthetically captured from a virtual point cloud and not optically captured from the real scene. Doing so, without repeating the real scene capturing step, one can change the characteristics of the integral image: the number of horizontal and vertical EIs, the amount of parallax and the field of view. All can be set modifying the simulated cameras array.

3.1. Microimages Generation Process

The process can be divided into five steps:

1. *Point Cloud creation.* The scene is captured and its depth information is extracted. A point cloud representing the scene is generated merging the RGB and depth information.
2. *EIs capturing.* The EIs are generated using a VPA. The number of virtual pinhole cameras in vertical (horizontal) direction is set equal to the number of pixels behind each microlens of the InI monitor in vertical (horizontal) direction. To capture EIs, the VPA is placed far away from the scene. A trade-off between the resolution of the EIs and the absence of black pixels depends on the position of the VPA. If the VPA is set too close to the point cloud, some information of the scene is lost and black pixels appear in the EIs (for the same reason as in [12,13]). On the other hand, as the VPA is moved further from the point cloud, black pixels' issue disappears, but the scene is captured in the EIs with lower resolution. Therefore, the position of the VPA is set empirically at the minimum distance from the point cloud that ensures the absence of black pixels. Then, this value is refined to set the reference plane's position, as explained in Section 3.2.

3. *Shifted cropping.* A portion of ($L \times V$) pixels of every EIs is cropped as in Figure 2. Shifting the cropped region with a constant step between adjacent EIs, allows setting the reference plane of the final image. This sets the portion of the 3D scene that will be reconstructed inside and outside the InI display, changing the depth sensation. More details on the parameters used in this step are given in Section 3.2.
4. *Resize.* The cropped EIs (sub-EIs) are resized to the spatial resolution of the InI monitor, that is, the number of microlenses of the MLA in horizontal and vertical direction.
5. *Transposition.* The pixels are resampled as in Figure 3, to convert the sub-EIs to the final microimages to project in the 3D InI monitor.

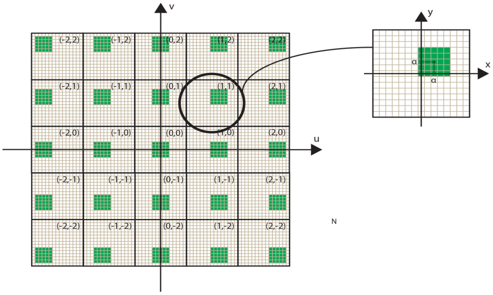


Figure 2. The EIs cropping. In the classical (u, v, x, y) parameterization of the integral image, let us say the central EI's coordinates are (u_0, v_0) and the central pixel's coordinates of every EI are (x_0, y_0). Considering the EI having coordinates (u, v), the central pixel of its sub-EI has coordinates (x, y) = ($x_0 + (u - u_0)\alpha$, $y_0 + (v - v_0)\alpha$).

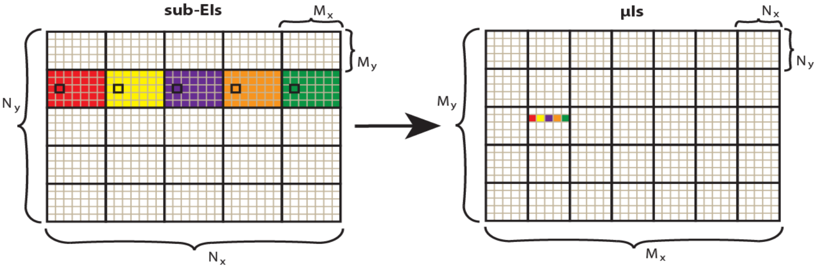


Figure 3. Transposition from sub-EIs to microimages. Starting from ($N_x \times N_y$) sub-EIs, each one with ($M_x \times M_y$) pixels, we obtain ($M_x \times M_y$) microimages each one with ($N_x \times N_y$) pixels. The correspondence is: $p_{i,j}(\mu I_{k,l}) = p_{k,l}(sub\text{-}EI_{i,j})$, where p means pixel.

3.2. Geometrical Model

One great feature of this algorithm is that, exploiting a geometrical model of the system, its parameters are adjusted to precisely set the position of the reference plane and the field of view of the final image.

To set the reference plane's position, the distance of the VPA and the shifting factor of the sub-EIs (α in Figure 2) can be adjusted. The procedure is the following. First, the VPA's position is set to the initial value vpa_i. As explained above, it has to be far enough from the scene to avoid black pixels' issue. Then, a provisional value of the shifting factor of the sub-EIs is calculated (β in Figure 4). As shown in Figure 4, it is the number of pixels of shifting which makes all the sub-EIs converge on the plane at $z = ref$

$$\beta = \frac{p \times g}{vpa_i - ref}$$

where p is the pitch between the cameras and g is the gap of the VPA. As the number of pixels of shifting must obviously be an integer number, the value of β must be rounded:

$$\alpha = round(\beta)$$

Now, if we consider these values of vpa and α, the reference plane would not be at $z = ref$. Therefore, we have to move the VPA's position to

$$vpa_f = \frac{p \times g}{\alpha} + ref$$

Operating on these two parameters (α and vpa), the reference plane's position can be set with very high resolution and precision. In [11], only α could be changed, because the distance of the cameras is fixed in the optical capturing stage. Thus, the image can be reconstructed only at determined planes, which depend on the physical and optical parameters of the real scene capture.

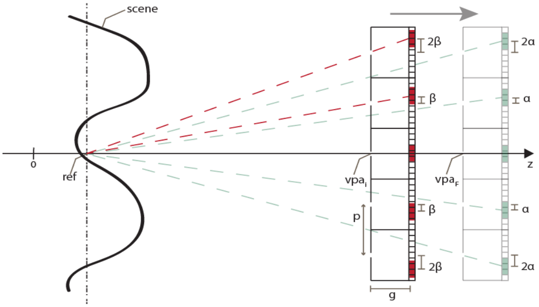


Figure 4. The procedure to set the reference plane's position. Note that, for graphical convenience, a VPA with pitch equal to the dimension of a single pinhole camera is shown. Actually, the pitch is smaller, so the virtual pinhole cameras overlap with each other.

The parameter that sets the field of view of the final image is the number of pixels of the sub-EIs. Let us assume, for instance, that we want to represent all the useful scene given by a point cloud having a width W. As we can see in Figure 5, the width of the cropped area has to be

$$L = \frac{g \times W}{d}$$

with d being the distance between the VPA and the reference plane. If we want to reduce the field of view to a $1/z$ portion of the scene, it is sufficient to divide by z the previous expression of L. The number of pixels in the other dimension (in vertical direction, if we assume that L is the number of pixels in horizontal direction) depends on the resolution of the InI display:

$$V = \frac{R_y}{R_x} L$$

with R_y and R_x being the vertical and horizontal spatial resolution of the InI display respectively. As explained in the previous section, the sub-EIs are finally resized from $(L \times V)$ to $(R_x \times R_y)$, before the final transposition.

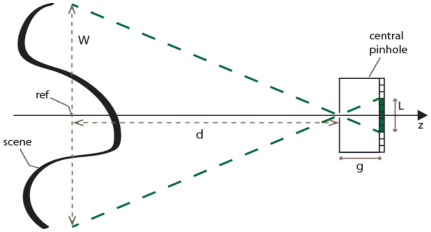


Figure 5. Setting the cropping factor. To set the number of pixels of the sub-EIs, a simple geometrical relation can be exploited: $L = \frac{g \times W}{d}$ (obviously, L has to be rounded to the nearest integer number). Reducing L, we can reduce the field of view.

4. Experimental Results

To demonstrate the effectiveness of the proposed method, we applied it to a real 3D scene that we captured with a Lytro Illum plenoptic camera. The experimental setup is shown in Figure 6a. The RGB image and the depth map (Figure 6b,c) are extracted through the Lytro Desktop software.

For the implementation of the 3D InI monitor, we used a Samsung SM-T700 tablet, with a screen of 14.1492 pixels/mm (ppm), and a MLA (Fresneltech, model 630) composed by lenslets of focal length f = 3.3 mm and pitch p = 1.0 mm. We have exactly 14.1492 pixels per microlens, thus, in the algorithm, we set the VPA to have 15 × 15 virtual pinhole cameras. In the resize stage, the sub-EIs are resized to 151 × 113 to make maximum use of the MLA in horizontal direction (the MLA is square-shaped, with a side of 151 mm, so it has 151 × 151 lenslets). Doing so, the final image will be 2265 × 1695. Nevertheless, it is important to remark the fact that the real number of pixels per microlens is 14.1492, so the image is finally rescaled to 2136 × 1599 (rescale factor k = 14.1492/15).

Figure 7 presents a comparison between the results obtained with the proposed technique and with the technique of [12], using the same point cloud as input. In the top row, the three images generated with the latter, with the reference plane set at three different depths are shown. In the bottom row, the images obtained with the proposed technique, with the reference plane set exactly at the same depths of the corresponding images of the top row. Clearly, in the images obtained with the concurrent method, a large black area (no information area) appears in the region close to the reference plane's position. Instead, in the images obtained with the proposed method, this problem does not occur. There are just some black pixels due to occlusions or bad depth estimation. Figure 8 shows the image of Figure 7d projected in our 3D InI display prototype. The reference plane is set at the background, so the map, the doll and the colored bow are reconstructed on the MLA plane, while the rest of the scene is reconstructed floating in front of it. Video S1 (https://youtu.be/X9UJGEh4CdE) is a video recording the InI display, which is much more effective than a single picture to perceive the 3D sensation of the reconstructed scene.

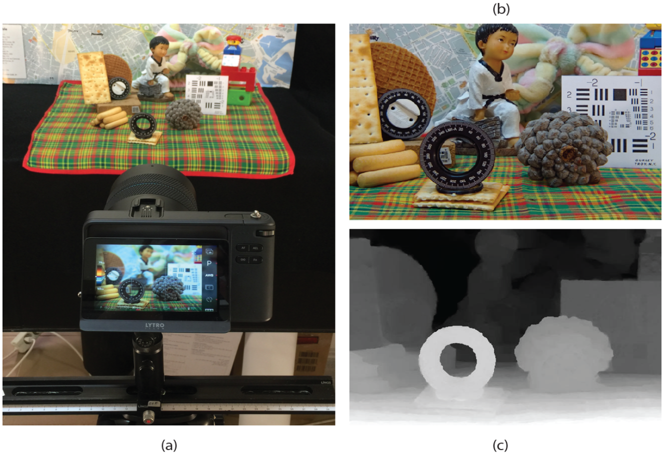


Figure 6. (**a**) Experimental set-up of the real capture of the 3D scene. (**b**) The RGB image extracted from Lytro Desktop software. (**c**) The depth map of the scene extracted from Lytro Desktop software.

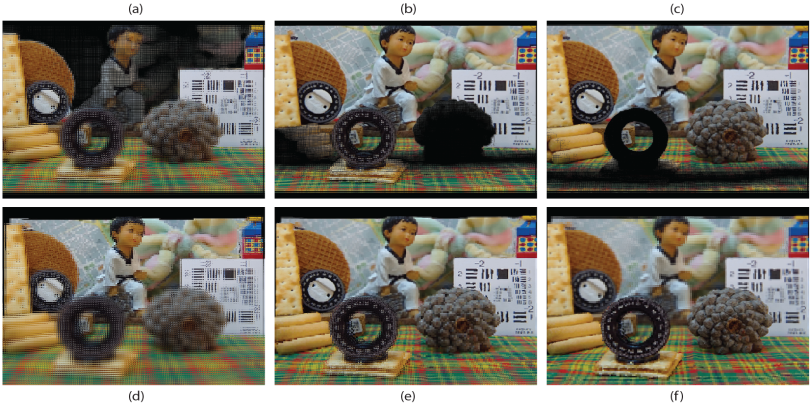


Figure 7. Top row: Images obtained with the technique of [12]. Bottom row: Images obtained with the proposed technique. (**a**,**d**) Reference plane at the background; (**b**,**e**) reference plane at the middle; and (**c**,**f**) reference plane at the foreground.

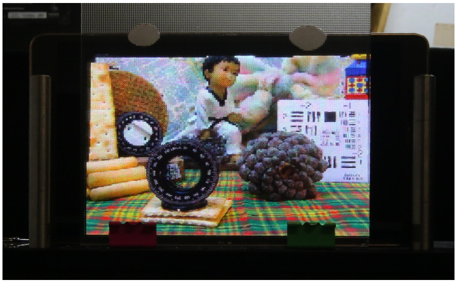


Figure 8. The integral image of Figure 7d projected in our 3D InI display prototype.

In Figure 9, the field of view is reduced by a factor $z = 2$. The image obtained is shown in Figure 9a, while in Figure 9b we show the image projected in the InI display.

In Figure 10, the pitch p of the pinhole cameras of the VPA is increased by a factor 5. The reference plane is set at the same position of the image of Figure 7d. Increasing the pitch has a double effect: the parallax amount increases while the depth of field (DOF) of the image decreases. On the contrary, setting an excessively low value of p increases the DOF, but the scene is reconstructed very flatly and with low parallax, thus losing the 3D sensation. The DOF of the image has to be adjusted to the DOF of the InI monitor to obtain a good reconstruction of the 3D scene. Here, p is intentionally set to a very high value in order to show the effect of an excessively large pitch. In Figure 10a, the integral image is shown. Comparing with Figure 7d, the DOF is greatly reduced: only the background map is in focus, while the rest of the objects appear increasingly defocused as we move away from the reference plane. In Figure 10b, the image projected in the InI monitor is shown. Note that the objects that are close to the reference plane are reconstructed well by the InI monitor, while the ones that are far from it appear really defocused. For matter of comparison, in Figure 8, which shows the image obtained with a fair value of p projected in the InI monitor, the whole scene is reconstructed well.

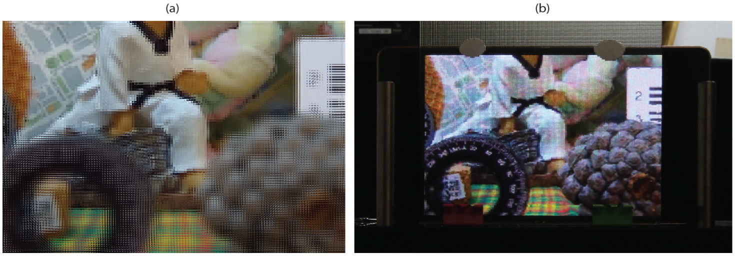


Figure 9. Field of view selection: (**a**) The integral image obtained with the same reference plane and pitch of Figure 7d, with half its field of view ($z = 2$). (**b**) Projection in 3D InI display.

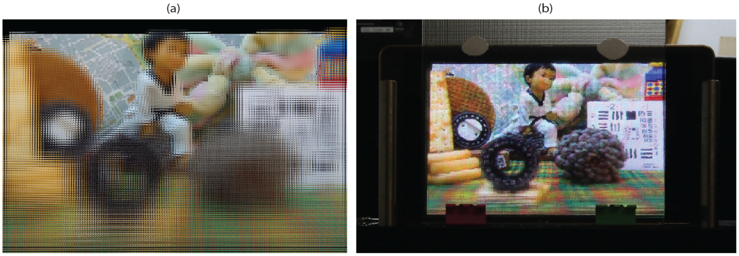


Figure 10. Pitch selection: (**a**) The integral image obtained with the same reference plane and field of view of Figure 7d, with five times its pitch. The DOF is greatly reduced with respect to the image of Figure 7d. (**b**) Projection in 3D InI display. The objects appear increasingly defocused as we move far from the reference plane.

5. Conclusions and Future Work

We have proposed a new method for the generation of microimages projectable in 3D InI displays. The method works with real 3D scenes and is adaptable to any InI display. It solves the problem of black pixels' areas and allows setting with precision the field of view of the final image and its reference plane, so controlling the way the scene is reconstructed by the InI monitor.

In the future work, the main focus will be on the real-time implementation of the system. Capturing the EIs through the VPA is the most time-consuming step, so it must be reassessed. Another goal is to resolve occlusions by using the information of multiple views of the integral image of the Lytro. The idea is to extract the depth maps of the lateral EIs and merge the RGB with the depth information of all these views into a single point cloud.

Author Contributions: Conceptualization, N.I., S.H., M.M.-C. and G.S.; Funding acquisition, M.M.-C. and G.S.; Investigation, N.I. and S.H.; Methodology, N.I.; Resources, M.M.-C. and G.S.; Software, N.I. and S.H.; Writing—original draft, N.I.; and Writing—review and editing, M.M.-C.

Funding: This work was supported by the Ministerio de Economía y Competitividad, Spain (Grant No. DPI 2015-66458-C2-1R), and by Generalitat Valenciana, Spain (project PROMETEOII/2014/072). S.H. acknowledges a predoctoral grant from University of Valencia (UV-INVPREDOC15-265754).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Son, J.Y.; Javidi, B. Three-dimensional imaging methods based on multiview images. *J. Disp. Technol.* **2005**, *1*, 125–140. [CrossRef]
2. Dodgson, N.A.; Moore, J.; Lang, S. Multi-view autostereoscopic 3D display. In *International Broadcasting Convention*; Citeseer: State College, PA, USA, 1999; Volume 2, pp. 497–502.
3. Lippmann, G. Epreuves reversibles donnant la sensation du relief. *J. Phys. Theor. Appl.* **1908**, *7*, 821–825. [CrossRef]
4. Raytrix. 3D Lightfield Camera. Available online: https://raytrix.de/ (accessed on 24 August 2018).
5. Martínez-Corral, M.; Dorado, A.; Barreiro, J.C.; Saavedra, G.; Javidi, B. Recent advances in the capture and display of macroscopic and microscopic 3-D scenes by integral imaging. *Proc. IEEE* **2017**, *105*, 825–836. [CrossRef]
6. Kwon, K.C.; Park, C.; Erdenebat, M.U.; Jeong, J.S.; Choi, J.H.; Kim, N.; Park, J.H.; Lim, Y.T.; Yoo, K.H. High speed image space parallel processing for computer-generated integral imaging system. *Opt. Express* **2012**, *20*, 732–740. [CrossRef] [PubMed]

7. Jiao, S.; Wang, X.; Zhou, M.; Li, W.; Hong, T.; Nam, D.; Lee, J.H.; Wu, E.; Wang, H.; Kim, J.Y. Multiple ray cluster rendering for interactive integral imaging system. *Opt. Express* **2013**, *21*, 10070–10086. [CrossRef] [PubMed]
8. Li, S.L.; Wang, Q.H.; Xiong, Z.L.; Deng, H.; Ji, C.C. Multiple orthographic frustum combing for real-time computer-generated integral imaging system. *J. Disp. Technol.* **2014**, *10*, 704–709. [CrossRef]
9. Chen, G.; Ma, C.; Fan, Z.; Cui, X.; Liao, H. Real-time Lens based Rendering Algorithm for Super-multiview Integral Photography without Image Resampling. *IEEE Trans. Vis. Comput. Gragh.* **2018**, *24*, 2600–2609. [CrossRef] [PubMed]
10. Navarro, H.; Martínez-Cuenca, R.; Saavedra, G.; Martínez-Corral, M.; Javidi, B. 3D integral imaging display by smart pseudoscopic-to-orthoscopic conversion (SPOC). *Opt. Express* **2010**, *18*, 25573–25583. [CrossRef] [PubMed]
11. Martínez-Corral, M.; Dorado, A.; Navarro, H.; Saavedra, G.; Javidi, B. Three-dimensional display by smart pseudoscopic-to-orthoscopic conversion with tunable focus. *Appl. Opt.* **2014**, *53*, E19–E25. [CrossRef] [PubMed]
12. Hong, S.; Dorado, A.; Saavedra, G.; Martínez-Corral, M.; Shin, D.; Lee, B.G. Full-parallax 3D display from single-shot Kinect capture. In *Three-Dimensional Imaging, Visualization, and Display 2015*; International Society for Optics and Photonics: Bellingham, WA, USA, 2015; Volume 9495.
13. Hong, S.; Ansari, A.; Saavedra, G.; Martinez-Corral, M. Full-parallax 3D display from stereo-hybrid 3D camera system. *Opt. Lasers Eng.* **2018**, *103*, 46–54. [CrossRef]
14. Piao, Y.; Qu, H.; Zhang, M.; Cho, M. Three-dimensional integral imaging display system via off-axially distributed image sensing. *Opt. Lasers Eng.* **2016**, *85*, 18–23. [CrossRef]
15. Cho, M.; Shin, D. 3D integral imaging display using axially recorded multiple images. *J. Opt. Soc. Korea* **2013**, *17*, 410–414. [CrossRef]

Paper VI

GPU-accelerated integral imaging and full-parallax 3D display using stereo–plenoptic camera system

Seokmin Hong, Nicolò Incardona, Kotaro Inoue, Myungjin Cho,
Genaro Saavedra, and Manuel Martínez-Corral

in *Optics and Lasers in Engineering* 115, 172–8 (2019).

Optics and Lasers in Engineering 115 (2019) 172–178

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng

GPU-accelerated integral imaging and full-parallax 3D display using stereo–plenoptic camera system

Seokmin Hong[a,*], Nicolò Incardona[a], Kotaro Inoue[b], Myungjin Cho[b], Genaro Saavedra[a], Manuel Martinez-Corral[a]

[a] *3D Imaging and Display Laboratory, Department of Optics, University of Valencia, 46100 Burjassot, Spain*
[b] *Department of Electrical, Electronic and Control Engineering, IITC, Hankyong National University, Anseong, 17579, South Korea*

ARTICLE INFO

Keywords:
3D display
Integral imaging
3D data registration
Point cloud
GPU
Plenoptic camera
Stereo camera

ABSTRACT

In this paper, we propose a novel approach to produce integral images ready to be displayed onto an integral-imaging monitor. Our main contribution is the use of commercial plenoptic camera, which is arranged in a stereo configuration. Our proposed set-up is able to record the radiance, spatial and angular, information simultaneously in each different stereo position. We illustrate our contribution by composing the point cloud from a pair of captured plenoptic images, and generate an integral image from the properly registered 3D information. We have exploited the graphics processing unit (GPU) acceleration in order to enhance the integral-image computation speed and efficiency. We present our approach with imaging experiments that demonstrate the improved quality of integral image. After the projection of such integral image onto the proposed monitor, 3D scenes are displayed with full-parallax.

1. Introduction

During the last century, three-dimensional (3D) imaging techniques have been spotlighted due to their merit of recording and displaying 3D scenes. Among them, integral imaging (InI) has been considered as one of the most promising technologies. This concept was proposed first by G. Lippmann in 1908. He presented the possibility of capturing the 3D information and reconstructing the 3D scene by using an array of spherical diopters [1–3]. Nowadays, the pickup procedure is performed by placing an array of tiny lenses, which is called microlens array (MLA), in front of the two-dimensional (2D) imaging sensor (e.g. CCD, CMOS). A collection of microimages is obtained, which is referred to as integral image. Interestingly, every microimage contains the radiance (spatial and angular) information of the rays. This is because different pixels of one microimage correspond to different incidence angles of the rays passing through each paired microlens. Figs. 1 and 2 show the comparison between a conventional and an InI (also known as plenoptic of light-field) camera. Several companies announced their plenoptic camera, which is based on Lippmann's integral photography theory [4–6]. On the other hand, in the display stage the MLA is placed in front of a screen, where is projected the integral image. The microlenses integrate the rays proceeding from the pixels of the screen and thus, reconstruct the 3D scene. Consequently, when the integral image is projected onto an InI display, observers can see the 3D scene with full-parallax and quasi-continuous perspective view.

In the meanwhile, many research groups are investigating how to acquire the depth map from the plenoptic image [7–9]. Sabater et al. [7] modeled demultiplexing algorithm in order to compose a proper 4D Light-Field (LF) image, and calculate the disparities from a restored sub-images array by using block-matching algorithm. Huang et al. [8] built their stereo-matching algorithm, and utilized it into their own framework named Robust Pseudo Random Field (RPRF) to estimate the depth map from the plenoptic image. Jeon et al. [9] calculated the depth map from an array of sub-aperture images by using the derived cost volume, multi-label optimization propagates, and iterative refinement procedure. We mainly applied Jeon's approach in our experiment.

The main contribution of this paper is to utilize the stereo–plenoptic camera system in order to get dense depth map from a pair of captured plenoptic images and get rid of the constraints of monocular vision system. Normally, multiple views can enlarge the field of view and recover the occluded information by complementing each other. For this reason, we can restore the depthless areas of the scene. Another important benefit from our proposal is to yield nicer quality of the integral image using a registered pair of point clouds. Besides, the use of the GPU acceleration technique assists to enhance the integral image's generation speed.

[*] Corresponding author.
E-mail address: seokmin.hong@uv.es (S. Hong).

https://doi.org/10.1016/j.optlaseng.2018.11.023
Received 28 September 2018; Received in revised form 20 November 2018; Accepted 26 November 2018
Available online 3 December 2018

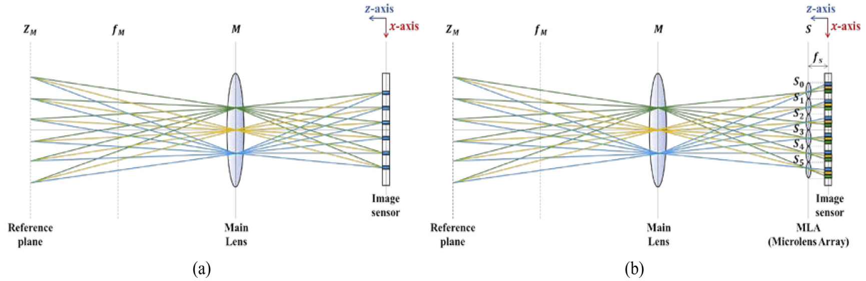


Fig 1. Scheme of image capturing system: (a) is a conventional camera; and (b) is a plenoptic camera. The pixels of (a) integrate, and therefore discard, the angular information even if they have. On the contrary, (b) can pick up both spatial and angular information thanks to the insertion of the microlens array.

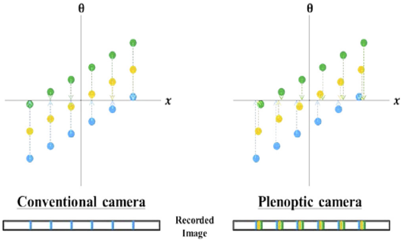


Fig 2. Illustration of the projected pixels to the imaging sensor, which are shown in the plenoptic field. The projected pixel from the conventional camera system gathers into a single pixel. However, plenoptic camera system projects all different incident information in independent pixel's position. This collected image becomes an integral image.

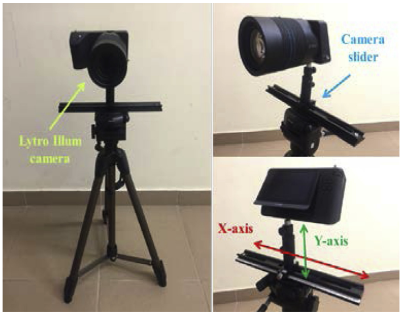


Fig 3. Proposed stereo–plenoptic camera system.

This paper is organized as follows. In Section 2., our previous related works are described. In Section 3., our contribution to compose and manage the point cloud from a pair of captured plenoptic images is illustrated. In Section 4. the methodology to generate an integral image from registered 3D information by using GPU acceleration technique is explained. Finally, in Sections 5. and 6. the experimental results are provided and the conclusions are carried out, respectively.

2. Related work

The closest work which is related with a stereo-type capturing and modification method has been published by our group recently. In [10] we exploited the stereo-hybrid 3D camera system composed of two Kinect sensors (Kinect v1 and v2), to take profit of different features for obtaining a denser depth map. Furthermore, we illustrated the benefit of binocular approach contrary to monocular one with some experimental results. However, the main distinction from current proposal is that [10] utilized hybrid camera set-up and obligatorily considered the remedy of the dissimilarities. Most of all, the working distance of the cameras used is restricted because of the usage of an infrared (IR) sensing technique. In this paper, we exploit the commercial plenoptic camera, named Lytro Illum. The important thing is that plenoptic cameras are passive devices in the sense that they do not need any additional light emitter. It can record the scene from the ambient light source directly. It means that the working distance of this camera is related to the camera lenses' optical properties. Furthermore, this plenoptic camera can decide the reference plane of the scene thanks to the InI's features [11,12].

In the meantime, [13] illustrated our approach to generate an integral image from a point cloud, which is ready to be projected onto an InI monitor. However, the bottleneck of this approach was that it required a long computational time. To solve this critical defect, in current approach we exploit GPU acceleration technique to generate microimages in parallel way, reducing the processing time.

3. Stereo–plenoptic image manipulation

In order to implement the stereo system, it is convenient to use two cameras of the same model. Accordingly, in our experimental system we utilized the camera slider in order to capture the scene in each different position with a single plenoptic camera, and we placed a tripod eager to configure the camera's proper position. Fig. 3 shows the camera setup and Fig. 4 shows the overview of our experimental environment. In Section 3.1, we describe our approach to manipulate the plenoptic image and obtain the depth map from this handled image. In sequence, in Section 3.2, we explain the methodology for the arrangement and registration process of a pair of point clouds.

3.1. Plenoptic image manipulation

Our proposal in this paper is the use of commercial plenoptic camera. Its software provides various functions: it helps to choose the proper perspective view, changes the focused plane of the scene, and extracts the calculated depth map (or disparity map), color image, and an encoded raw image format [5]. Fortunately, [14,15] help to decode the

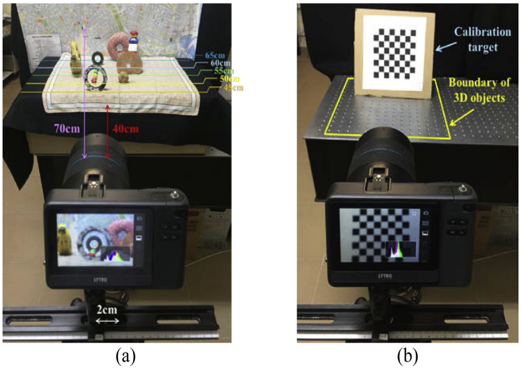


Fig 4. Overview of proposed experimental environment: (a) is capture for the main scene, and (b) is capture for the calibration process.

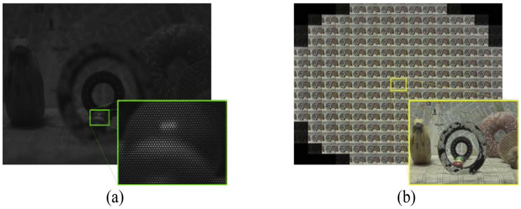


Fig 5. (a) is a raw plenoptic image from plenoptic camera, (b) is a composed sub-aperture image array from plenoptic image. See text for further details.

raw plenoptic image and extract sub-images from this encrypted data. Interestingly, this extracted raw data contains a grayscale image (see Fig. 5(a)). The main reason is that there is a Bayer color filter array over the camera's sensor to capture the colors. Thus, it must be demosaiced to get the color information back. It is noticeable that the color tones of captured images shown in Fig. 4(a) and 5(b) are different. Note, however, that the first is the image extracted from Lytro software and the other is a sub-image extracted through [14,15]. The main reason of that difference is that they use different Bayer demosaicing algorithms. We extract the sub-aperture images array in order to follow [9] approach (see Fig. 5(b)), which estimates the depth map by minimizing stereo-matching costs between sub-images with sub-pixel accuracy, and corrects the unexpected distortions. However, even after correcting the distortion problem via the referenced algorithm, the estimated depth map still has some image distortion effect. Thus, we performed the plenoptic camera calibration and rectification before the depth map calculation. The diagram of Fig. 6 shows our approach well.

Fig. 7 shows the comparison between our proposed depth map estimation strategy and the output from Lytro's software (Lytro Destktop v.5.0.1). Fig. 7(a, b) have more continuous depth levels and stable gradation than Fig. 7(c, d). On the contrary, the sharpness of the targets and the shape of the object's surfaces in the former are worse than in the latter.

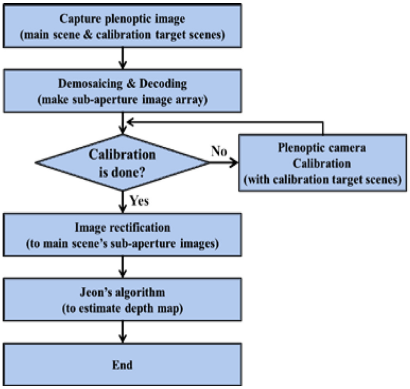

Capture plenoptic image
(main scene & calibration target scenes)
Demosaicing & Decoding
(make sub-aperture image array)
Calibration is done?
No
Plenoptic camera Calibration
(with calibration target scenes)
Yes
Image rectification
(to main scene's sub-aperture images)
Jeon's algorithm
(to estimate depth map)
End

Fig 6. Flow chart of our proposed depth estimation strategy.

3.2. Point cloud modification and registration

The aim of this section is to explain how to compose the point cloud from the image, and to make registration from one point cloud to the other in order to arrange them in a proper position. In [13], we composed the point cloud from a pair of color and depth map images. We assigned six values to each point of the point cloud, namely its (x, y, z) coordinates and RGB color intensities. Each point of the RGB image corresponds to the point of the depth image having the same (x, y) coordinates. So it is sufficient to assign the corresponding depth value to all the points of the RGB image. Finally, this modified 3D information is arranged into the virtual 3D space. Afterward, we need to make registration between left and right point clouds. This is because the two scenes are mutually shifted and it is necessary to arrange them in a proper way. To solve this issue, we utilize Iterative-Closest-Point algorithm (ICP), as in [10]. ICP calculates the movement and minimizes the distance be-

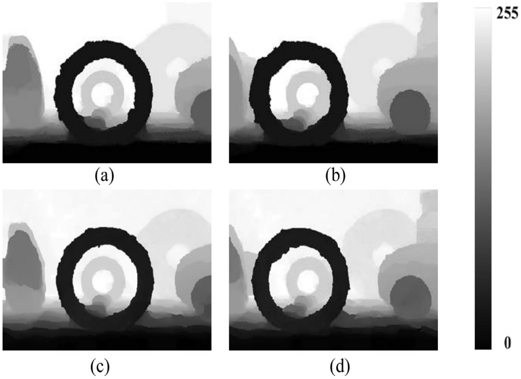


Fig 7. The depth map comparison result: top row images (a, b) are the estimated depth map result from our approach, while bottom row images (c, d) are from the output of Lytro's software. Right bar shows the depth intensity value from depth map image (0: closest area, 255: farthest area).

tween point clouds. As is well known, ICP is often used to reconstruct 2D or 3D data captured from different positions. The output of ICP algorithm is a rigid-body transformation matrix, which includes translation and rotation information [16–18]. This matrix permits to refer the position of one point cloud to the other in appropriate way.

4. Integral image generation from the point cloud with GPU acceleration

Once aligned the pair of point clouds, the resulting one is ready to generate an integral image. As we mentioned in [13], the production of an integral image is processed in a virtual 3D space using a virtual pinhole array (VPA). We place the VPA in a proper position in the virtual 3D scene, and all the points from the cloud are projected through all the pinholes by using back-projection technique, as in [19]. Interestingly, the location of the VPA will represent the position of the displayed image's reference plane. For instance, a point located behind the VPA will be reconstructed behind the MLA, while a point in front of the VPA will be reconstructed floating in front of it. Each point projected through the pinholes forms the microimages' pixels and finally, this entire back-projection mapping calculation produces the integral image.

On the other hand, we also need to consider the scale factor between input image and integral image's sizes. The main reason is that the scale factor decides the nearest-neighbor interpolation's index, as in [13]. This interpolation helps to fill the empty pixels during the back-projection mapping and as a result, proper interpolation index helps to improve the quality of the integral image. Eq. 1 and 2 show how to derive scale factors:

$$\begin{cases} Dst_w = II_w \\ Dst_h = \frac{II_w}{Org_w} \times Org_h \end{cases} \tag{1}$$

$$\lambda_{u,v} = \frac{Dst_{w,h}}{Org_{w,h}} \tag{2}$$

Where II_w is target integral image's width size, $Org_{w,\,h}$ is input image, $Dst_{w,\,h}$ is final integral image size, and $\lambda_{u,\,v}$ is scale factor, respectively.

However, these back-projection mapping and interpolation processes are heavy work. In order to solve this drawback, we utilize the GPU acceleration technique. The use of central processing units (CPUs) computation has the limitation due to their general purpose of usage. Even if CPUs have their own threads to compute, their performance is not sufficient to boost the computation speed because of the way of CPU's sequential implementation process and the limited number of CPU Cores (the number of threads depends on the capacity of CPU's Cores). On the contrary, GPU computation enables to execute thousands of threads to compute their mission in parallel [20,21]. It means that we can compute the integral image in a parallel way and as a result, we can speed up the computation time. Fig. 8 shows our approach and the comparison scheme between CPU and GPU computation. After this process we can get the integral image, which is ready to be displayed in an InI monitor.

5. Experimental results

In our experiment, we register the right point cloud into the space of the left one. The main reason is that the right scene not only contains the occluded information of the left scene, but also new objects appear. On the other hand, regarding the display part, we utilized the Samsung SM-T700 (14.1338px/mm) tablet as screen, and we mounted a MLA which has focal length $f = 3.3$ mm and pitch $p = 1.0$ mm (Model 630 from Fresnel Technology). We utilized 152 × 113 microlenses from this MLA because this is the maximum possible usage for the screen used (see Fig. 10's InI monitor set-up). A noteworthy feature is that the number of pinholes of the VPA must match the number of microlenses. The generated microimage is composed of 15 × 15 pixels, and thus, the full size of the integral image is 2280 × 1695 pixels. Finally, we need to resize the integral image to take into account the real number of pixels per microlens, so the image is finally resized to 2148 × 1597 (resizing factor $k = 14.1338$px/15px). Fig. 9 shows the result of produced integral images.

To show our experimental result, we composed the set-up as shown in Fig. 10. Originally, our main target are binocular observers, who can see the 3D nature of displayed scene. Unfortunately, the full-parallax effect cannot be directly demonstrated in a manuscript or even in a monocular video. In order to demonstrate this 3D effect, we replaced the binocular observer with a monocular digital camera, as recording device. A collection of pictures is obtained displacing the camera in horizontal and vertical direction. Media 1 and 2 show the result obtained with each left and right scenes, and Media 3 shows the result of the proposed method. Fig. 11 shows this experimental result with more details. Our proposed result has better quality than each, left and right, captured scenes. For instance, left and right scenes have black areas (depthless areas) caused by occlusions. On the other hand, our proposed method restores these occluded areas thanks to the registration and complementation between left and right captured scenes.

Meanwhile, we exploit the parallelism in integral image computation via NVIDIA CUDA programming model, which is a software platform for solving non-graphics problems in a parallel way [21]. Our hardware specification is the following: Intel i7 4cores in CPU, and NVIDIA

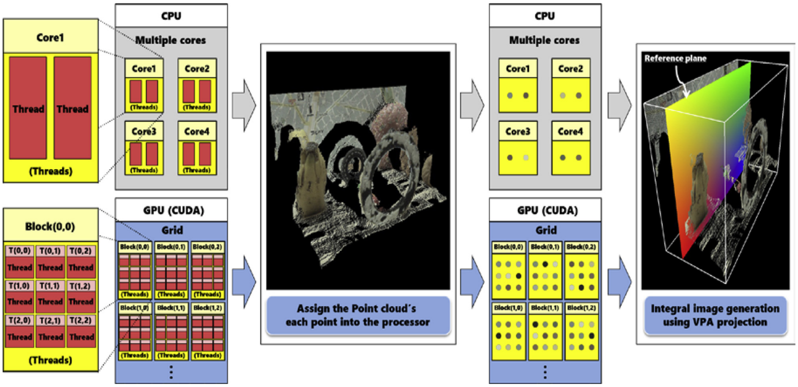


Fig 8. The comparison scheme how to compose an integral image in CPU and GPU computation. Each thread picks single 3D point from the point cloud and computes the proper pixels of an integral image using VPA projection. From the third step, GPU is able to assign thousands of points in a same time contrary to CPU.

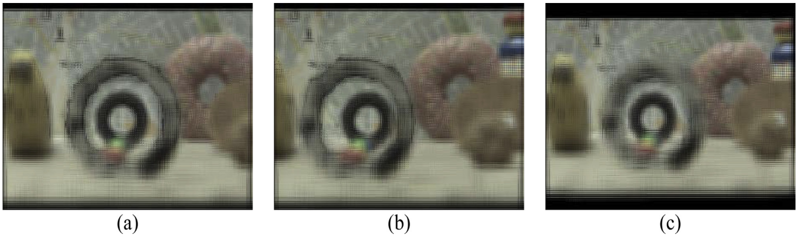


Fig 9. Composed integral image: (a) is from left scene, (b) is from right scene, and (c) image is registered scene between left and right scenes.

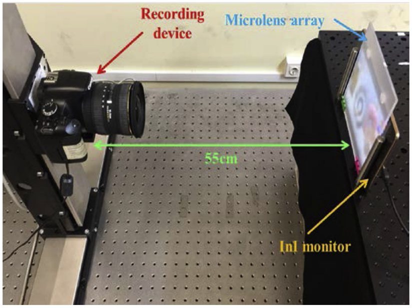


Fig 10. Overview of experimental system.

Table 1
More detail of comparison result between CPU and GPU computation time.

List	CPU(Sec.)		GPU(Sec.)	
Kind of the scene	Left, right	Registered	Left, right	Registered
0 interpolation	109.71	224.59	29.59	60.57
1 interpolation	302.39	629.87	30.47	63.46
2 interpolations	699.40	1432.56	32.68	66.94
3 interpolations	1281.99	2610.55	53.77	109.48

GeForce GTX 870 M in GPU. We tested the algorithm with various interpolation indices to compare the computation speeds (see Fig 12 and Table 1). We have found that the GPU implementation is much faster than CPU, especially when we increase the interpolation index. In fact, the interpolation index does not affect the computation time in the GPU implementation.

6. Summary and conclusion

In this paper we utilized the stereo–plenoptic camera system to display the captured plenoptic image into an InI monitor and enhance the quality of the displayed 3D image. We did a plenoptic camera calibration and rectification to solve the tilted and distorted plenoptic image's defect. Furthermore, we extracted the sub-aperture images array from the calibrated plenoptic image in order to estimate the depth map. This calculated depth map is used to compose the 3D point cloud, which is arranged into the virtual 3D space. Then we performed a registra-

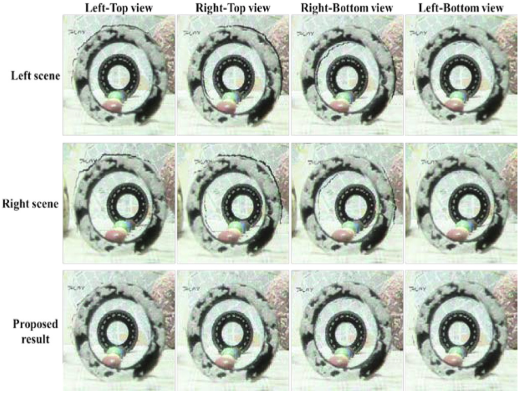


Fig 11. Comparison result between displayed integral images: first row is from left scene, second row is from right scene, and third row is our proposed result. All the images are excerpted from recorded video (Media 1, 2, and 3), and we clipped-out a specific part at the scene in order to emphasize the comparison result clearly.

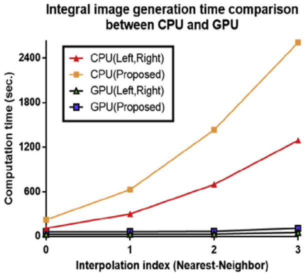


Fig 12. The integral image generation time comparison between CPU and GPU. The triangle represents the left and right scene's result, and the rectangle represents the registered scene's result.

tion between left and right scene's point clouds to arrange them in a proper position. This fused point cloud has denser 3D data and manages to recover the depthless areas properly. Finally, we generated the integral image via VPA through the back-projection method. To boost the computation time, we adopted GPU acceleration technique in this procedure. This generated integral image is displayed in our proposed integral imaging monitor and it displays an immersive scene with full parallax to the binocular observers.

In the future work, the main focus will be on the real-time implementation of the system using different and/or newer types of 3D cameras: stereo-vision camera [22,23], or even higher quality of plenoptic camera [4,6]. Another goal is to enhance the accuracy of 3D data registration using non-rigid objects mapping [24–26].

Acknowledgement

This work was supported by the Ministerio de Economía y Competitividad, Spain (Grant No. DPI 2015-66458-C2-1R), by Generalitat Valenciana, Spain (project PROMETEOII/2014/072), and by the National Research Foundation of Korea (NRF-2017R1D1A1B03030343, NRF-2017K1A3A1A19070753). S. Hong acknowledges a Predoctoral contract from University of Valencia (UV-INVPREDOC15-265754).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.optlaseng.2018.11.023.

References

[1] Lippmann G. Epreuves réversibles photographies integrals. *Comptes Rendus de l'Académie des Sciences* 146 1908:446–51.
[2] Lippmann G. Epreuves reversibles donnant la sensation du relief. Journal de Physique Théorique et Appliquée 1908;7:821–5.
[3] Lippmann G. L'étalon international de radium. Radium (Paris) 1912;9:169–70.
[4] . Raytrix camera http://www.raytrix.de.
[5] . Lytro camera https://www.lytro.com.
[6] . PiCam: Pelican Imaging Camera http://www.pelicanimaging.com.
[7] Sabater N, Seifi M, Drazic V, Sandri G, Perez P. Accurate disparity estimation for plenoptic images. In: Proceedings European Conference on Computer Vision; 2014. p. 548–60.
[8] Huang C-T. Robust pseudo random fields for light-field stereo matching. IEEE Conf Comput Vis Pattern Recognit 2017:11–19.
[9] Jeon H, Park J, Choe G, Park J, Bok Y, Tai Y, Kweon I. Accurate depth map estimation from a lenslet light field camera. IEEE Conf Comput Vis Pattern Recognit 2015:1547–55.
[10] Hong S, Ansari A, Saavedra G, Martinez-Corral M. Full-parallax 3D display from stereo-hybrid 3D camera system. Opt Laser Eng 2018;103:46–54.
[11] Martinez-Corral M, Dorado A, Navarro H, Saavedra G, Javidi B. Three-dimensional display by smart pseudoscopic-to-orthoscopic conversion with tunable focus. Appl Opt 2014;53:19–25.
[12] Scrofani G, Sola-Pikabea J, Llavador A, Sanchez-Ortiga E, Barreiro JC, Saavedra G, Garcia-Sucerquia J, Martinez-Corral M. FIMic: design for ultimate 3D-integral microscopy of in-vivo biological samples. Biomed *Opt Express* 2018;9:335–46.
[13] Hong S, Shin D, Lee B, Dorado A, Saavedra G, Martinez-Corral M. Towards 3D television through fusion of kinect and integral-imaging concepts. J Disp Technol 2015;11:894–9.
[14] Dansereau DG, Pizarro O, Williams SB. Decoding, calibration and rectification for lenselet-based plenoptic cameras. IEEE Conf Comput Vis Pattern Recognit 2013:1027–34.
[15] Dansereau DG. Lightfield toolbox for matlab http://dgd.vision/Tools/LFToolbox.
[16] Besl PJ, Mckay ND. A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell 1992;14:239–56.
[17] Zhang Z. Iterative point matching for registration of free-form curves and surfaces. Int J Comput Vis 1994;13:119–52.
[18] Rusinkiewicz S, Levoy M. Efficient variants of the ICP algorithm. In: Proceedings Third Int. Conf. on 3-D Digit. Imaging and Mode; 2001. p. 145–52.
[19] Martínez-Corral M, Javidi B, Martínez-Cuenca R, Saavedra G. Formation of real, orthoscopic integral images by smart pixel mapping. Opt Express 2005;13:9175–80.
[20] "NVIDIA CUDA Toolkit, version 9.1", https://developer.nvidia.com/cuda-toolkit.
[21] Bilgic B, Horn BKP, Masaki I. Efficient integral image computation on the GPU. IEEE Int Intell Veh Symp 2010:528–33.
[22] Bumblebee2 .
[23] *ZED* .
[24] Hähnel D, Thrun S, Burgard W. An extension of the ICP algorithm for modeling nonrigid objects with mobile robots. Int Joint Conf Artif Intell 2003.
[25] Chui H, Rangarajan A. A new point matching algorithm for non-rigid registration. Comput Vis Image Underst 2003;89:2–3.

[26] Ma J, Zhao J, Jiang J, Zhou H. Non-rigid point set registration with robust transformation estimation under manifold regularization. Proc Conf Artif Intell 2017:4218–24.

Seokmin Hong received the B.Eng. and M.Sc. degrees in digital and visual contents from Dongseo University, Busan, South Korea, in 2012 and 2014, respectively. In 2012, Dongseo University honored him with the B.Eng. Extraordinary Award. Since 2015, he has been working with the 3D Imaging and Display Laboratory, Optics Department, University of Valencia, Spain. His research interests are image processing, computer vision, 3D display, 3D integral imaging, and applied computer science.

Nicolò Incardona received the B.Eng. and M.Sc. degrees in Electronic Engineering from Polytechnic University of Milan, Italy, in 2013 and 2016 respectively. He developed his master's thesis at Polytechnic University of Valencia, Spain. Since 2017 he has been working with the 3D Imaging and Display Laboratory, University of Valencia, Spain. His research interests are image processing, 3D display and integral microscopy.

Kotaro Inoue received the B.S. and M.S. degrees in computer science and electronics from Kyushu Institute of Technology, Fukuoka, Japan, in 2015 and 2017, respectively. He is currently a doctoral student at Hankyong National University in Korea. His research interests include visual feedback control, 3D display, 3D reconstruction, and 3D integral imaging.

Myungjin Cho received the B.S. and M.S. degrees in Telecommunication Engineering from Pukyong National University, Pusan, Korea, in 2003 and 2005, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Connecticut, Storrs, CT, USA, in 2010 and 2011, respectively. Currently, he is an associate professor at Hankyong National University in Korea. He worked as a researcher at Samsung Electronics in Korea, from 2005 to 2007. His research interests include 3D display, 3D signal processing, 3D biomedical imaging, 3D photon counting imaging, 3D information security, 3D object tracking, 3D underwater imaging, and 3D visualization of objects under inclement weather conditions.

Genaro Saavedra received the B.Sc. and Ph.D. degrees in physics from Universitat de València, Spain, in 1990 and 1996, respectively. His Ph. D. work was honored with the Ph.D. Extraordinary Award. He is currently Full Professor with Universitat de València, Spain. Since 1999, he has been working with the “3D Display and Imaging Laboratory”, at the Optics Department. His current research interests are optical diffraction, integral imaging, 3D high-resolution optical microscopy and phase-space representation of scalar optical fields. He has published on these topics about 50 technical articles in major journals and 3 chapters in scientific books. He has published over 50 conference proceedings, including 10 invited presentations.

Manuel Martinez-Corral was born in Spain in 1962. He received Ph. D. degree in Physics in 1993 from the University of Valencia, which honored him with the Ph.D. Extraordinary Award. He is currently Full Professor of Optics at the University of Valencia, where he co-leads the “3D Imaging and Display Laboratory”. His teaching experience includes lectures and supervision of laboratory on Geometrical Optics, Optical Instrumentation, Diffractive Optics and Image Formation for undergraduate and Ph.D. students. Fellow of the SPIE since 2010 and Fellow of the OSA since 2016, his research interest includes microscopic and macroscopic 3D imaging and display technologies. He has supervised on these topics fifteen Ph. D. students (three of them honored with the Ph.D. Extraordinary Award), published over 120 technical articles in major journals (which have received more than 2.700 citations), and pronounced over fifty invited and keynote presentations in international meetings. He is co-chair of the Three-Dimensional Imaging, Visualization, and Display Conference within the SPIE meeting in Defense, Security, and Sensing. He is Topical Editor of the OSA journal Applied Optics.