Immersive 3D Holoscopic Video System

Amar Aggoun, Emmanuel Tsekleves, and Mohammad Rafiq Swash
Brunel University, UK

Dimitrios Zarpalas, Anastasios Dimou, and Petros Daras
Informatics and Telematics Institute, Greece

Paulo Nunes and Luís Ducla Soares
Instituto de Telecomunicações and Instituto Universitário de Lisboa (ISCTE-IUL), Portugal

A 3D imaging system utilizes 3D holoscopic imaging to display full-color images with continuous parallax within a wide viewing zone.

Recent film releases such as *Avatar* have revolutionized cinema by combining 3D technology and content production and real actors, leading to the creation of a new genre at the outset of the 2010s. The success of 3D cinema has led several major consumer electronics manufacturers to launch 3D-capable televisions and broadcasters to offer 3D content. Today's 3DTV technology is based on stereo vision, which presents left- and right-eye images through temporal or spatial multiplexing to viewers wearing a pair of glasses. The next step in 3DTV development will likely be a multiview autostereoscopic imaging system, which will record and present many pairs of video signals on a display and will not require viewers to wear glasses.[1,2]

Although researchers have proposed several autostereoscopic displays, the resolution and viewing position is still limited. Furthermore, stereo and multiview technologies rely on the brain to fuse the two disparate images to create the 3D effect. As a result, such systems tend to cause eye strain, fatigue, and headaches after prolonged viewing because users are required to focus on the screen plane (accommodation) but to converge their eyes to a point in space in a different plane (convergence), producing unnatural viewing. Recent advances in digital technology have eliminated some of these human factors, but some intrinsic eye fatigue will always exist with stereoscopic 3D technology.[3]

These facts have motivated researchers to seek alternative means for capturing true 3D content, most notably *holography* and *holoscopic imaging*. Due to the interference of the coherent light fields required to record holograms, their use is still limited and mostly confined to research laboratories. Holoscopic imaging (also referred to as integral imaging) in its simplest form, on the other hand, consists of a lens array mated to a digital sensor with each lens capturing perspective views of the scene.[4–9] In this case, the light field does not need to be coherent, so holoscopic color images can be obtained with full parallax. This conveniently lets us adopt more conventional live capture and display procedures. Furthermore, 3D holoscopic imaging offers fatigue-free viewing to more than one person, independent of the viewers' positions.

Due to recent advances in theory and microlens manufacturing, 3D holoscopic imaging is becoming a practical, prospective 3D display technology and is thus attracting much interest in the 3D area. The 3D Live Immerse Video-Audio Interactive Multimedia (3D Vivant, www.3dvivant.eu) project, funded by the EU-FP7 ICT-4-1.5–Networked Media and 3D Internet, has proposed advances in 3D holoscopic imaging technology for the capture, representation, processing, and display of 3D holoscopic content that overcome most of the aforementioned restrictions faced by traditional 3D technologies. This article presents our work as part of the 3D Vivant project.

3D Holoscopic Content Generation

The 3D holoscopic imaging technique creates and represents a true volume spatial optical model of the object scene in the form of a planar intensity distribution using unique optical components. A 3D holoscopic image is recorded using a regularly spaced array of lenslets closely packed together with a recording device (see Figure 1a).[4] Each lenslet views the scene at a slightly different angle than its neighbor, so a scene is captured from many viewpoints and parallax information is recorded. We then replay the 3D holoscopic images by placing a microlens array on top of

1070-986X/13/$31.00 © 2013 IEEE Published by the IEEE Computer Society

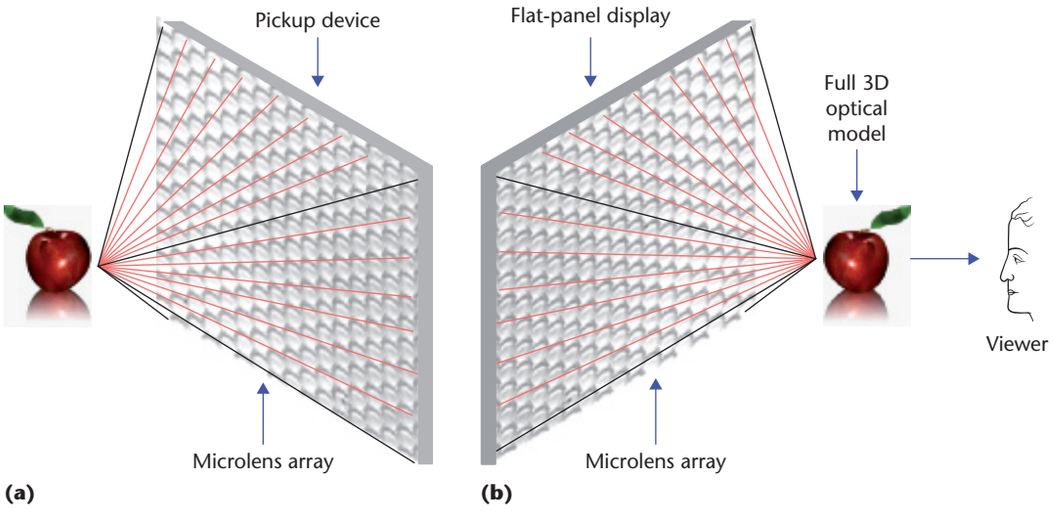


Figure 1. Recording and replay content using a 3D holoscopic imaging system. (a) Capture. (b) Replay.

the recoded planar intensity distributions that is illuminated by white light diffused from the rear. Figure 1b shows that the object is then constructed in space by the intersection of ray bundles emanating from each of the lenslets. In replay, the reconstructed image is pseudoscopic (inverted in depth). In the last two decades, researchers have proposed various optical and digital techniques to convert the pseudoscopic image to an orthoscopic image.[5–9]

The camera setup in Figure 1a does not possess depth control, so the reconstructed object appears in its original location in space. Therefore, with this setup, we can only produce 3D virtual images or 3D real pseudocsopic images. Furthermore, objects far from a microlens array will suffer from poor spatial sampling on sensor pixels. Therefore, this type of camera is best suited for close imaging applications. In addition, standard microlens manufacturing methods, such as UV and hot embossing, produce shrinkage and replication errors that can yield errors in the pitch of ±20 μm/20 mm (35 μm/35 mm), which is a particularly large percentage error for microlens arrays with a small pitch.

To remedy these two problems, the setup in Figure 2 adds objective and relay lenses. An objective lens provides depth control, which allows the image plane to be near the microlens array. The 3D holoscopic image's spatial sampling is determined by the number of lenses, so we can obtain higher-resolution images by reducing the lenses' size. A tradeoff exists between the number of lenses and the number of viewpoint images or pixels under each lens. These pixels define discrete locations in the

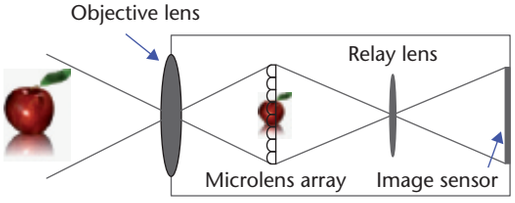


Figure 2. 3D holoscopic camera with objective and relay lenses. This camera setup remedies the problems introduced by the setup in Figure 1 by adding objective and relay lenses.

aperture of the objective lens from where the object can be viewed. Therefore, making the lenses smaller reduces the angular information about points in the object because fewer pixels can view it.

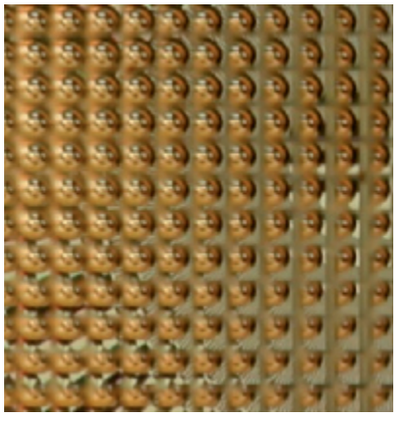
Live images are recorded in a regular block pixel pattern. The planar intensity distribution representing a 3D holoscopic image consists of a 2D array of $M \times M$ microimages, due to the structure of the microlens array used in the capture. A significant problem with the setup in Figure 2 is the vignetting of microimages at the microlens pickup stage due to the use of an image relay. We can resolve this by introducing a field lens at the back of the microlens array. Figures 3a and Figures 3b show images obtained without and with the field lens. The purpose of a field lens is to relay the pupil image to prevent vignetting of off-axis rays when using a system with an intermediate image plane. The field lens focal length is determined by the exit pupil and relay entrance pupil conjugates as measured from the microlens array image plane. As Figure 3b shows,

(a)

(b)

Figure 3. Holoscopic images captured (a) without and (b) with a field lens. The field lens relays the pupil image to prevent vignetting of off-axis rays when using a system with an intermediate image plane.

(a)



(b)

Figure 4. 3D holoscopic image. (a) Recorded using a 90-μm pitch microlens array. (b) Magnified section.

using the field lens removes most of the vignetting.

To achieve a fill factor above 95 percent, we can fit a square aperture on the front of the camera. Figure 4 illustrates that this, along with the regular structure of the square microlenses array used in the square grid (recording the microlens array), generates a regular structure in the intensity distribution.

3D Holoscopic Content Visualization

One of the aims of the 3D Vivant project is to investigate the possibility of displaying 3D holoscopic content on commercially available autostereoscopic displays. Currently, a number of autostereoscopic displays available on the market use a combination of lenticular optics along with liquid crystal display (LCD) panels. The lenticular elements are positioned at an angle to the LCD pixel array.[10] This mixes adjacent views, reducing image-flipping problems and spreading the effect of the black mask, making it less visible. The other benefit of this design is that each view has a better aspect ratio; instead of splitting the display horizontally into many views, both horizontal and vertical directions are split. The slanted lenticular arrangement requires subpixel mapping to image all pixels along a slanted line in the same direction.

To adapt 3D holoscopic content to autostereoscopic displays, we must use unidirectional 3D holoscopic images. These images are obtained using special 3D holoscopic imaging systems that use 1D cylindrical microlens array for capture and replay instead of a 2D array of microlenses. The resulting images contain parallax only in the horizontal direction and can be displayed using a lenticular sheet on an LCD panel. We would then use a 3D holoscopic virtual camera to generate eight orthographic views and a software tool to perform the subpixel mapping and interlacing of the orthographic views into a single image to be fed into the 3D autostereoscopic display. Figure 5 shows a photograph of a real hand in front of the screen showing a computer-generated 3D holoscopic image.

Coding and Transmission

To make 3D holoscopic video delivery feasible over real-world networks and storage media,

video codecs must deal efficiently with the large amount of data captured with high-resolution sensors. Consequently, new efficient video coding tools will become of paramount importance.

Due to the small angular disparity between adjacent microlenses, a significant cross-correlation exists between neighboring microimages. Therefore, this inherent cross-correlation of 3D holoscopic images can be seen as a type of self-similarity, and we can exploit it to improve coding efficiency. For example, Figure 6 shows two 3D holoscopic video frames with a different amount of self-similarity (SS) redundancy.

An earlier work proposed a scheme for SS estimation and compensation to explore the high SS between neighboring microimages and improve H.264/AVC performance.[11] To predict a block area, the SS estimation process uses block-based matching in a picture area that has been already coded and reconstructed to find a best match in terms of a suitable matching criterion (such as the sum of absolute differences). This prediction process for 3D holoscopic content is efficient because it does not need to know the precise structure of the underlying microlens array and, consequently, the arrangement of the microimages.

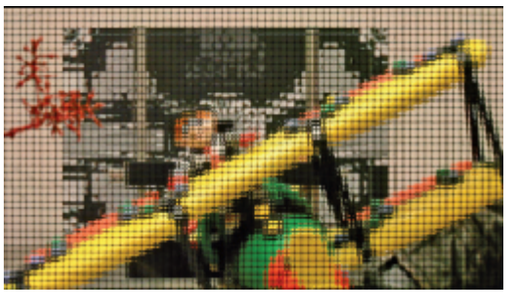
Recently, MPEG and the International Telecommunication Union-Telecommunication Standardization Sector (ITU-T) formed the Joint Collaborative Team on Video Coding (JCT-VC) to propose a standardization project on video coding called High Efficiency Video Coding (HEVC).[12] This new-generation video coding standard promises to improve the coding efficiency of the state-of-art H.264/AVC to fulfill the current compression efficiency requirements, among others, for high- and ultra-high-resolution video content. However, HEVC does not currently efficiently handle 3D holoscopic content because it does not take into account its inherent SS. We can overcome this by combining the flexible coding tools from HEVC with the SS estimation concept[11,13] by adding new predictive coding modes that exploit the special arrangement of 3D holoscopic content. An earlier work did this by defining two new prediction modes for HEVC as SS and SS-skip prediction modes, which take advantage of the inherent structure of 3D holoscopic images.[13]

Figure 5. Two perspective views of a 3D holoscopic image on a commercial 3D autostereoscopic display. Parts of the 3D holoscopic image on the screen (such as the first globe) are visible at the same depth as the real hand.

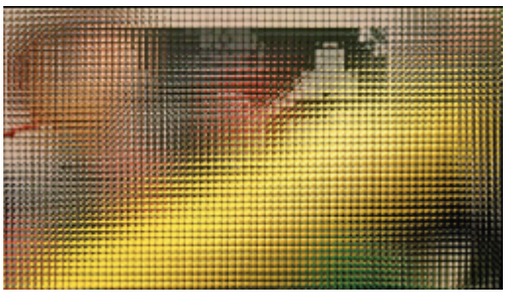
The SS prediction mode, which is based on the HEVC interprediction mode, enables all the original prediction unit (PU) partition patterns, but it replaces the motion estimation and compensation with the SS estimation and compensation process. When this mode is used, the relative position between a given PU and its prediction is encoded and transmitted as a vector (similar to a motion vector). This is referred to as the SS vector.[11,13] In addition, the prediction residual is also encoded and transmitted.

In SS estimation, we adapt the search range to the size of the PU being considered. In the SS compensation process, one self-similarity vector is derived for each PU. In addition, we use an adapted scheme based on the advanced motion vector prediction (AVMP) to select a SS vector prediction.

On the other hand, when the SS-skip prediction mode is used, only a SS vector is encoded and transmitted for each PU—that is, no



(a)



(b)

Figure 6. 3D holoscopic plane and toy test sequence captured using a 250-μm pitch microlens array. (a) Frame with less self-similarity (SS) redundancy. (b) Frame with more SS redundancy.

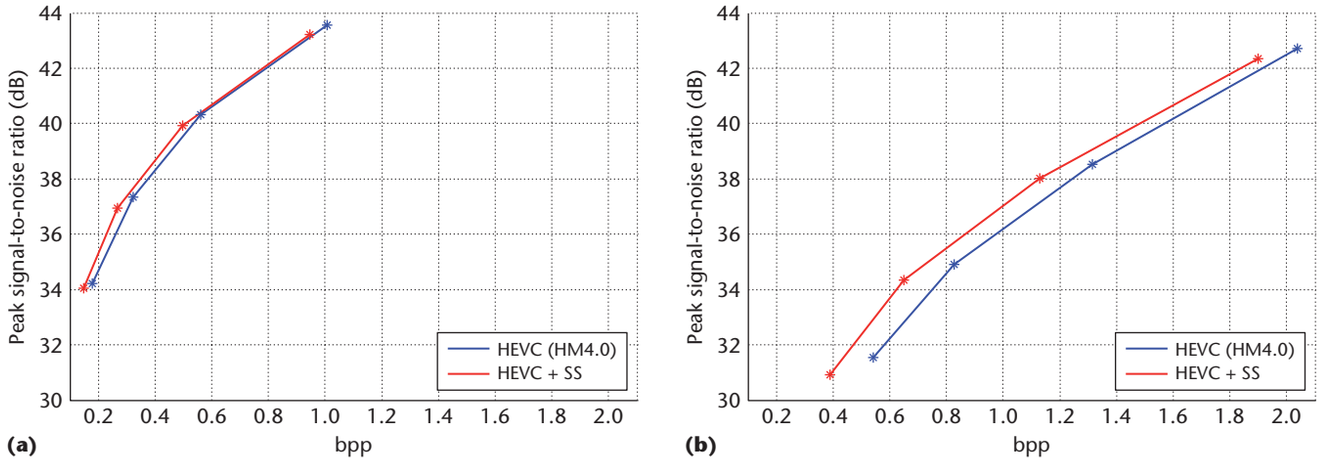


Figure 7. Peak signal-to-noise ratio (PSNR) results for the plane and toy test sequence. (a) Frame with less SS redundancy. (b) Frame with more SS redundancy.

residual data is sent to the decoder. This mode infers an SS vector for the current PU by considering SS vector candidates from spatially neighboring PUs. These inferred SS vectors are constrained to point to the area defined in the SS estimation process because this guarantees that the prediction indicated by the chosen SS vector (which belongs to the same frame) is already available as an SS reference.

Applying this coding scheme to 3D holoscopic content, such as the plane and toy sequence illustrated in Figure 6, shows that we can significantly improve the HEVC performance for this type of content (see Figure 7). Moreover, the improvements are highly dependent on the amount of SS redundancy existing in the video frames, as Figures 7a and Figures 7b illustrate. As such, the stronger the SS redundancy, the better the performance of the SS coding approach.

Content Interactivity

With 3D holoscopic video, users can see different perspectives of the same 3D scene by simply changing their position. However, for a truly interactive viewing experience, users should be able to individually select the 3D objects in the 3D holoscopic video scene and hyperlink to other content by clicking on them through search and retrieval of the desired 3D content. This section discusses 3D holoscopic object extraction, segmentation, and search and retrieval methods.

Depth Estimation

As a result of novel depth-extraction techniques, researchers have increasingly looked at the compactness of using 3D holoscopic imaging in depth measurement.[14] For example, Dimitrios Zarpalas and his colleagues presented a depth-computation method that involves extraction and matching a set of feature points, called *anchor points*, and then trying to fit a surface to the reconstructed features to optimally connect them.[15] The authors derived their framework from two earlier techniques.[16] As such, it proposes one surface-fitting optimization that is implemented through graph cuts and constrained by the anchor points. The framework first extracts anchor points, which are the 3D points of the strong correspondences among numerous successive viewpoint images. In every set of seven successive viewpoint images, one of them is regarded as the central one and is matched with the rest. If the stereo pairs of the same pixel from the central viewpoint agree on the same depth value with the majority of the rest images, then this estimated 3D point is regarded as reliable and is considered an anchor point.

The next step uses anchor points to constrain the optimization procedure, which otherwise can easily get stuck to local extrema due to the optimization's high complexity. This reduces the optimization complexity and enhances the depth estimation accuracy. In the proposed formulation, graph cuts help determine the optimized solution of the depth of the 3D points that correspond to each pixel, instead of the disparities between pixels of adjacent viewpoint images. The 3D scene is scanned uniformly along the emanating projection rays of the integral image. Furthermore, the proposed framework helps us model the piecewise nature of depth by introducing a twofold regularization term among adjacent pixels on both the

elemental and viewpoint images. This twofold neighborhood handling lets us reconstruct scenes with high spatial smoothness. Figure 8 depicts the proposed algorithm's workflow.

Figure 9 illustrates the results of the proposed algorithm for synthetic 3D holoscopic images. The estimated depth map is shown to contrast the scene's actual depth map, both of which were estimated from the central's viewpoint angle. In all cases, the different objects are correctly perceived and differentiated from their neighbors based on their estimated depth values. The reliability of the anchor point depth estimations is verified by their mean relative error, which is approximately 3.98 percent for the synthetic dataset (see ftp://ftp.iti.gr/pub/Holoscopy). The mean relative error for the complete image depth estimates is 6.13 percent for the whole synthetic dataset. Figure 10 shows results for real 3D holoscopic images.

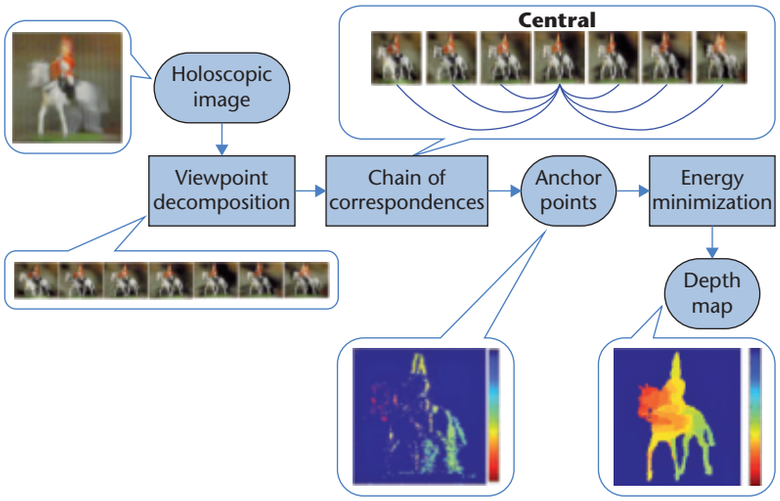


Figure 8. Schematic representation of the proposed procedure for depth estimation. This starts with decomposing the integral image onto its viewpoint images, which are used to gather the anchor points. Depth estimation is formulated as an optimization problem and anchor points are used to constrain it.

Search and Retrieval

To effectively browse collections of 3D holoscopic images, we created a search and retrieval framework by exploiting the rich information contained in them. Due to its nature, each holoscopic image is analyzed into three distinct visual modalities: the 2D

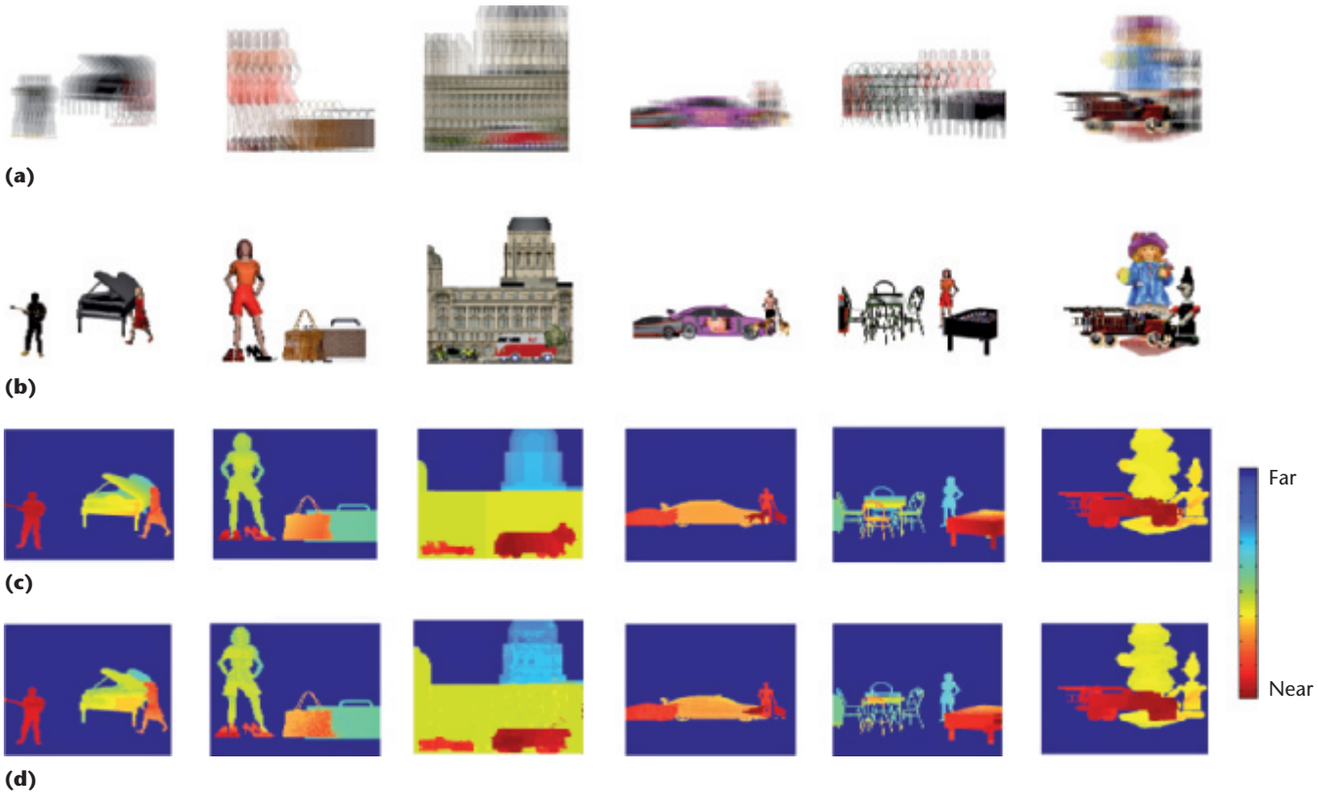


Figure 9. Results of the proposed algorithm on synthetic holoscopic images. (a) Holoscopic images. (b) Corresponding viewpoint images. (c) Actual depth map of the scene. (d) Depth map estimation using the proposed method.

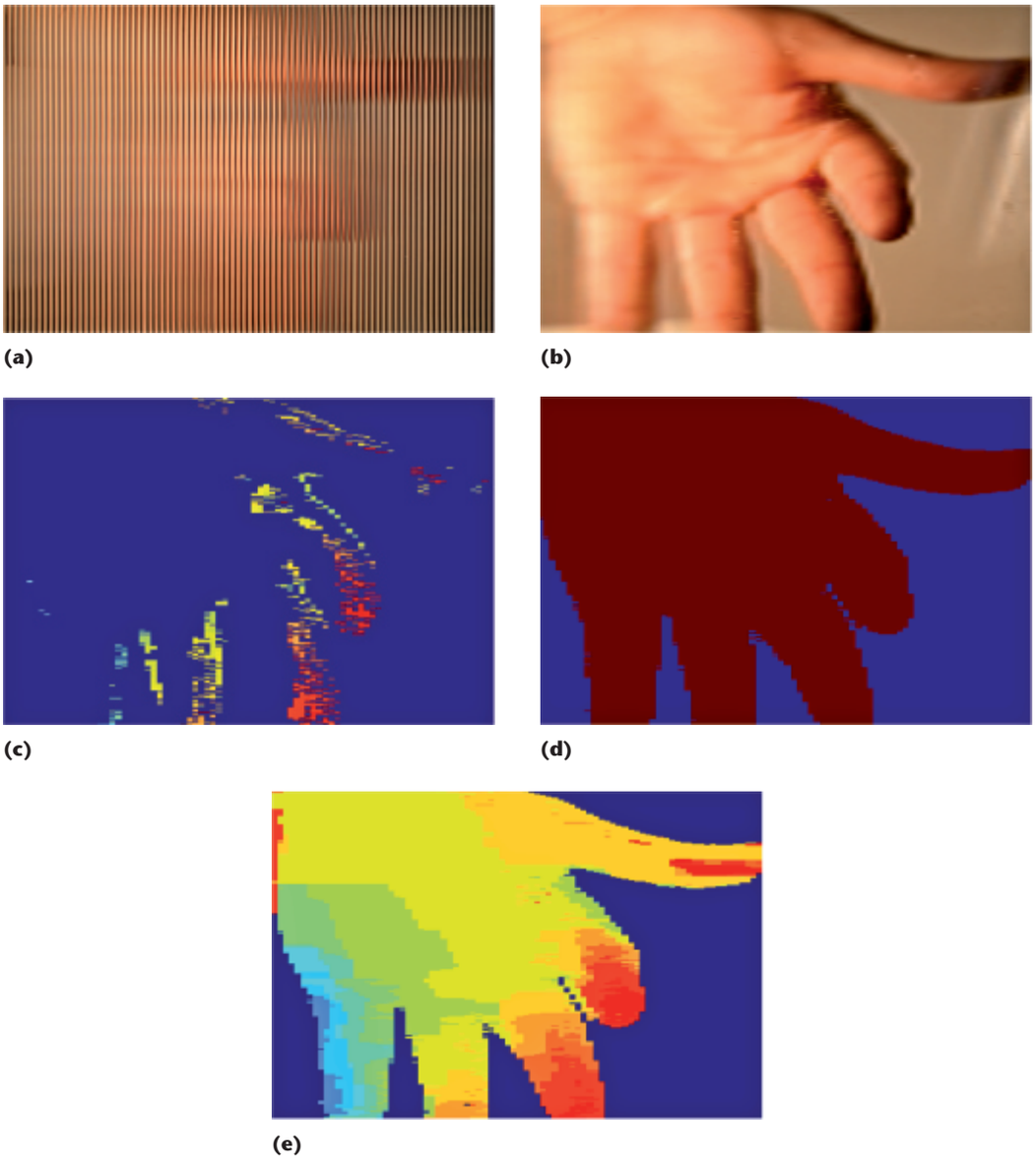


Figure 10. Results of a real holoscopic image captured with cylindrical lenses depicting a human palm. (a) Holoscopic image. (b) One viewpoint image. (c) Anchor points. (d) Foreground mask. (e) Extracted depth map using the proposed method.

viewpoint images, the depth information either as a depth map or as a 2.5D point cloud, and the 3D curvature information extracted from the depth map, again as a 2D curvature map.

For each of the modalities, we extract different state-of-the-art descriptors to choose the optimal combination. These descriptors are applied on the viewpoint images, depth maps, and curvature maps. Furthermore, on the 2.5D point cloud depth modality, we also use a novel local shape descriptor for 2.5D images, called *projection images*.[17] For a given point, the projection distances of the projected neighboring points are captured and separated into two sets (positive and negative projection images), based on which side of the projection plane they lie. The plane is divided into sectors by defining angular and radial divisions and the mean values of the positive and the negative projection distances for each sector are calculated. This way, we capture a compact, descriptive twofold representation of the object's shape.

We compute numerous local descriptors for each holoscopic image per modality and use the bag-of-words methodology to create a single descriptor vector for every image. We use a number of randomly selected local descriptor vectors to create clusters described as ''visual words'' using the K-Means algorithm. The histogram depicting the appearance frequency of

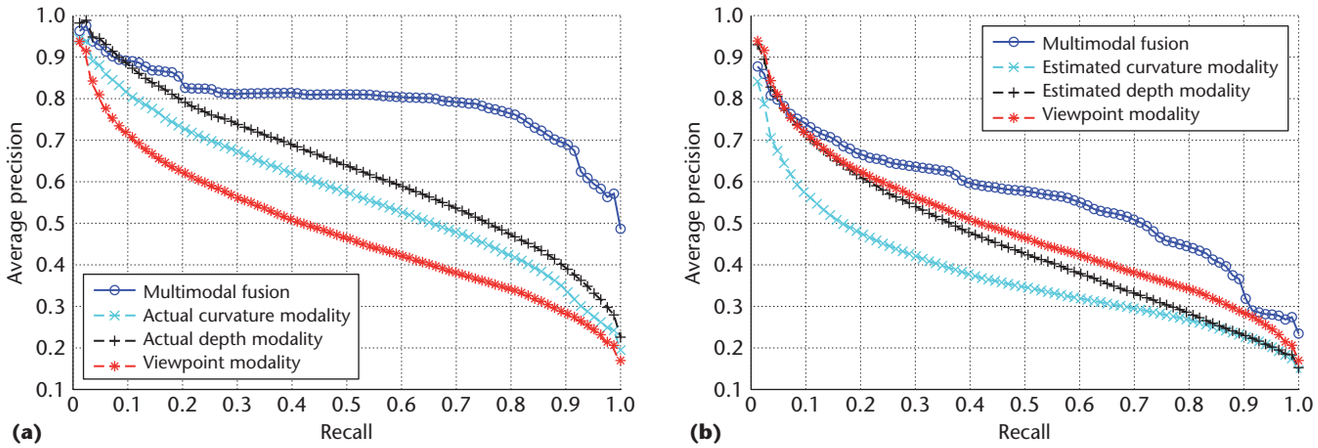


Figure 11. Precision recall for all modalities and the proposed framework. (a) Ground truth (actual) depth versus (b) estimated depth.

those visual words in each image constitutes the final descriptor vector.

To overcome the lack of spatial information associated with local descriptors, we use spatial pyramidal decomposition.[18] A range of distance metrics is also tested to identify the most suitable for each descriptor type. Then, for each modality, the different descriptors are combined through weighting each of them by optimizing—using the graph-cuts technique—the retrieval performance on a training dataset.

After combining descriptor vectors for each modality, the next step involves the fusion of the different modalities into a multimodal framework by combining all monomodal distance matrices to produce an over-all-modalities distance matrix. The latter is achieved by using a manifold-learning approach that relies on Laplacian eigenmaps utilizing a heat kernel to construct a multimodal low-dimensional feature space. In this space, each data point represents a holoscopic image. Such a framework lets us link the holoscopic content with other types of content (such as 2D photographic images, 3D models, and 2.5D point clouds).

Experimental Validation

To validate the proposed framework, we created a database containing 665 synthetic images with nine different classes—animals, beds, buildings, cars, chairs, couches, people, plants, and tables—containing synthetic holoscopic images consisting of their viewpoint images, along with estimates and ground truth of depth and curvature images.

We assessed each of the modalities using precision-recall curves and mean average precision (MAP). Figure 11 and Table 1 depict the corresponding results.

Table 1. Mean average precision for all modalities and the proposed multimodal fusing framework, using ground truth and estimated depth.

Modality	Actual depth MAP	Estimated depth MAP
mmFusion	0.7936	0.5573
Curvature	0.5722	0.3744
Depth	0.6325	0.4488
Viewpoint	0.4820	0.4820

The actual depth-value results show that the depth-related modalities are more significant than the texture-based modalities. Furthermore, the multimodal fusion framework strongly boosts the retrieval performance, proving that 3D holoscopic images should be regarded as multimodal content. The estimated depth results show that the retrieval performance depends on the depth-estimation quality.

Conclusion

We demonstrated a 3D holoscopic video system for 3DTV application. We showed that using a field lens and a square aperture significantly reduces the vignetting problem associated with a relay system and achieves over 95 percent fill factor. The main problem for such a relay system is the nonlinear distortion during the 3D image capturing, which can seriously affect the reconstruction process for a 3D display. The nonlinear distortion mainly includes lens radial distortion (intrinsic) and microlens array perspective distortion (extrinsic). This is the task of future work.

Our results also show that the SS coding approach performs better than the standard

HEVC scheme. Furthermore, we show that search and retrieval performance relies on the depth map's quality and that the multimodal fusion boosts the retrieval performance. **MM**

Acknowledgments

We acknowledge the support of the European Commission under the Seventh Framework Programme (FP7) project 3D Vivant (Live Immerse Video-Audio Interactive Multimedia).

References

1. Y. Zhu and T. Zhen, ''3D Multi-View Autostereoscopic Display and Its Key Technologies,'' *Proc. Asia-Pacific Conf. Image Processing* (APCIP 2009), vol. 2, IEEE CS, 2009, pp. 31–35.
2. G. Lawton, ''3D Displays Without Glasses: Coming to a Screen Near You Computer,'' *Computer,* vol. 44, no. 1, 2011, pp. 17–19.
3. L. Onural, ''Television in 3-D: What Are the Prospects,'' *Proc. IEEE,* vol. 95, no. 6, 2007, pp. 1143–1145.
4. G. Lippmann, ''Epreuves Reversibles Donnant la Sensation du Relief'' [Reversible Events Giving Sensation of Relief], *J. Physique Théorique et Appliquée,* vol. 7, no. 1, 1908, pp. 821–825.
5. A. Aggoun, ''3D Holoscopic Imaging Technology for Real-Time Volume Processing and Display,'' *High-Quality Visual Experience: Signals and Comm. Technology,* Springer, 2010, pp. 411–428.
6. J.S. Jang and B. Javidi, ''Formation of Orthoscopic Three Dimensional Real Images in Direct Pickup One Step Integral Imaging,'' *Optical Eng.,* vol. 42, no. 7, 2003, pp. 1869–1870.
7. F. Okano et al., ''Real-Time Pickup Method for a Three-Dimensional Image Based on Integral Photography,'' *Applied Optics,* vol. 36, no. 7, 1997, pp. 1598–1604.
8. M. Martínez-Corral et al., ''Formation of Real, Orthoscopic Integral Images by Smart Pixel Mapping,'' *Optics Express,* vol. 13, no. 23, 2005, pp. 9175–9180.
9. B. Javidi et. al., ''Orthoscopic, Long-Focal-Depth Integral Imaging by Hybrid Method,'' *Proc. SPIE,* vol. 6392, SPIE, 2006, pp. 639203-1–639203-8.
10. C. van Berkel, D.W. Parker, and A.R. Franklin, ''Multiview 3D LCD,'' *Proc SPIE,* vol. 2653, SPIE, 1996; doi:10.1117/12.237437.
11. C. Conti et al., ''Spatial Prediction Based on Self-Similarity for Compensation for 3D Holoscopic Image and Video Coding,'' *Proc. IEEE Int'l Conf. Image Processing* (ICIP), IEEE, 2011, pp. 961–964.
12. K. Ugur et al., ''High Performance, Low Complexity Video Coding and the Emerging HEVC Standard,'' *IEEE Trans. Circuits and Systems for Video Technology,* vol. 20, no. 12, 2010, pp. 1688–1697.
13. C. Conti, P. Nunes, and L.D. Soares, ''New HEVC Prediction Modes for 3D Holoscopic Video Coding,'' *Proc. IEEE Int'l Conf. Image Processing* (ICIP), IEEE, 2012.
14. C. Wu et al., ''Depth Map from Unidirectional Integral Images Using a Hybrid Disparity Analysis Algorithm'' *IEEE J. Display Technology,* vol. 4, no. 1, 2008, pp. 101–108.
15. D. Zarpalas et al., ''Depth Estimation in Integral Images by Anchoring Optimization Techniques,'' *Proc. IEEE Int'l Conf. Multimedia & Expo* (ICME), IEEE CS, 2011; doi:10.1109/ICME.2011.6011887.
16. D. Zarpalas et al., ''Anchoring-Graph-Cuts towards Accurate Depth Estimation in Integral Images,'' *IEEE J. Display Technology,* vol. 8, no. 7, 2012, pp. 405–417.
17. D. Zarpalas, G. Kordelas, and P. Daras, ''Recognizing 3D Objects in Cluttered Scenes Using Projection Images,'' *Proc. 8th IEEE Int'l Conf. Image Processing* (ICIP 2011), IEEE, 2011, pp. 673–676.
18. S. Lazebnik, C. Schmid, and J. Ponce, ''Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories,'' *Proc. IEEE Conf. Computer Vision & Pattern Recognition* (CVPR), IEEE CS, 2006, pp. 2169–2178.

Amar Aggoun is a reader in information and communication technologies at Brunel University, UK. His research interests include light-field imaging systems, computer generation and live capture of 3D integral images, depth measurement and volumetric data reconstruction, 3D medical visualization, 3D video coding, computer vision systems, and real-time digital image/video processing. Aggoun has a PhD in electronic engineering from the University of Nottingham, UK. Contact him at amar.aggoun@Brunel.ac.uk.

Emmanuel Tsekleves is a lecturer in multimedia design and technology at Brunel University, UK. His research interests include virtual reality, 3D applications and games for education, and health and defense applications. Tsekleves has a PhD in electronic and computer engineering from Brunel University London, UK. Contact him at Emmanuel.Tsekleves@Brunel.ac.uk.

Mohammad Rafiq Swash is a research assistant and doctoral student in 3D holoscopic imaging

technologies at Brunel University, UK. Swash graduated with a first-class honors degree in computer system engineering from Brunel University. Contact him at Mohammad.swash@brunel.ac.uk.

Dimitrios Zarpalas is an associate researcher at the Informatics and Telematics Institute, Greece. His research interests include 3D medical image processing, shape analysis, 3D object recognition, and search and retrieval and classification of 3D objects. Zarpalas has an MSc in computer vision from Pennsylvania State University. Contact him at zarpalas@iti.gr.

Anastasios Dimou is an associate researcher at the Information Technologies Institute, Greece. His research interests include image and video processing and analysis, Semantic Web Technology, video coding, smart surveillance systems, and biomedical applications. Dimou has an PDEng in information and communication technology the Eindhoven University of Technology, The Netherlands. Contact him at dimou@iti.gr.

Petros Daras is a researcher (grade C) at the Informatics and Telematics Institute, Greece. His research interests include processing, retrieval, and recognition of 3D objects; medical informatics; medical image processing; and bioinformatics. Daras has a PhD in electrical and computer engineer from the Aristotle University of Thessaloniki, Greece. Contact him at daras@iti.gr.

Paulo Nunes is an assistant professor at the Instituto de Telecomunicações and Instituto Universitário de Lisboa (ISCTE-IUL), Portugal. His research interests include 3D video processing and coding. Nunes has a PhD in electrical and computer engineering from the Instituto Superior Técnico, Portugal. He is a member of IEEE. Contact him at paulo.nunes@lx.it.pt.

Luís Ducla Soares is an assistant professor at the Instituto de Telecomunicações and Instituto Universitário de Lisboa (ISCTE-IUL), Portugal. His research interests include 3D video processing and coding. Soares has a PhD in electrical and computer engineering from the Instituto Superior Técnico, Portugal. He is a member of IEEE. Contact him at lds@lx.it.pt.