B 1.2 Progress beyond the state of the art

We review here the current state of the art and current research in capture, image and video-based technologies. *Capture* refers to the first step, which is taking enough and appropriate images and videos to calibrate the cameras (i.e., determine the 3D position and orientation of cameras in the world) and typically reconstruct an approximate geometric representation of the scene. This can be seen in the figure below. *Image based* techniques use the input images (blue views in the figure) to create a novel view, from a new camera position which was not in the input data (red in the figure). A notoriously hard problem with these methods is changing the lighting conditions: in contrast to traditional digital assets, lighting is included in the photograph and video, so the difficult problem of separating light from the underlying intrinsic materials has to be solved first before we can actually change the lighting.

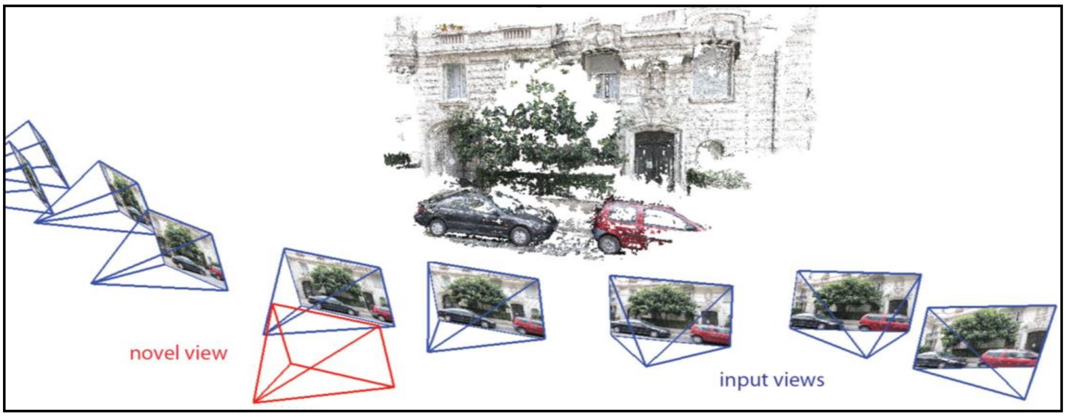

Fig 1.10: Several cameras are used to capture a scene; First input camera (blue) positions and orientations are recovered (calibrated) and approximate geometry computed. We can then create novel views (red) using Image-based rendering.

Finally, *Video-Based rendering* refers to technologies which allow the capture of repetitive and/or stochastic phenomena such as trees moving in the wind, or crackling flames of a fire in video, and the ability to also place them in the 3D scene, akin to the way cameras are placed in the IBR process in the figure above.
We review the related literature below. We also briefly discuss related, but different approaches, namely procedural modelling methods and “Google Maps Games”, and position ourselves with respect to them.

All references marked in **BOLD** are from consortium partners, illustrating the top-level expertise of our consortium in this domain.

B1.2.1 Capture

Capture of imagery for playback or browsing purposes is an increasingly popular research topic in both computer vision and graphics. [Snavely et al. 2006] showed an attractive IBR-like interface for browsing collections of photos taken at tourist landmarks, where sparse 3D reconstruction allowed the images to be rendered as billboards that could be navigated with simplistic spatial controls. That technology has now matured into a mobile-phone app from Microsoft called “Photosynth,” and the current state of the art IBR rendering system from our team **[Goesele et al. 2010]**. While not providing capture-planning guidance to the user (which we do), the recent work of [Davis et al. 2012] showed an interface that gives users feedback while capturing raw imagery for IBR purposes.

With editing applications in mind, technological progress for visual capture has largely focused on 3D geometry of simple constrained indoor scenes. Without the IBR methods we propose here, the quest for good and cheap geometry capture is an essential and still-elusive objective of computer vision and graphics. [Bouguet and Perona 1998] presented one of the earliest light-weight capture system, which promised to allow home users to acquire approximate depth maps of medium-sized objects “on the desktop.” Most recently, the KinectFusion work of [Newcombe et al. 2011] has delivered on that promise by leveraging the Microsoft Kinect structured-light camera and a carefully designed user interface, so novice users really are capable of capturing coarse-scale geometry of, for example, a small room interior. Such methods however do not generalize to outdoors environments, and require a somewhat involved hardware setup.

For moderate detail on small objects indoors or outdoors, though still assuming they are matte painted, the work by [Ummenhofer and Brox 2012] achieves dense multi-view reconstruction by coupling a surface reprojection energy with their camera bundle adjustment optimization. This and related dense-stereo techniques, i.e. multi-view with small inter-camera separations, assume that surfaces are highly textured and not shiny, and because they rely on resolving the correspondence problem, really struggle to capture fine surface details like the bumps on a building, on a leaf, or on skin **[Goesele et al.2007]**.

To capture fine-scale surface details, correspondence cues must be either replaced or supplemented with photometric response cues, i.e. that a curvy or bumpy surface reveals its shape, even when untextured, because sections look brighter or darker, in accordance with how much the surface normal is pointing toward a light source. Since the pioneering work of [Woodham 1980] established the technique of photometric-stereo based on this premise, researchers in our community have been working to overcome the limitations imposed by that technique’s assumptions, namely that the surface is stationary with respect to the camera, that the surface has a single Lambertian reflectance, that the camera has a linear response curve, that there is only one light source at a time, that this light source is infinitely far away, and that the surface is imaged under multiple (usually three or more) lighting configurations.

For example, [Hertzmann and Seitz 2005] can handle materials with any reflectance properties, as long as a calibration reference object for each material is available, while [Alldrin et al. 2008] cope with a greater variety of material properties by first estimating a set of basis reflectance functions. Subsequent to [Zhang et al. 2003], which combined Structure-from-Motion, multi-view stereo, and photometric stereo (the last two having been combined by [Nehab et al. 2005]), [Joshi and Kriegman 2007] were able to obtain a single consistent highly-detailed normal map, despite seeing the sample object at various orientations. Combining multi-view stereo with photometric stereo for textureless shiny objects was explored in [Hernandez et al. 2008]. Closest to our proposed approach, [Higo et al. 2009] is the most advanced in the thread of hybrid multi-view and photometric stereo techniques. While they allow the light source to move and come close to the object (i.e. no longer expecting it to shine as a point source at infinity), they still only capture desktop-sized mostly matte toy scenes indoors.

One interesting avenue for capturing highly specular surfaces has explored the use of structured light patterns. Systems like [Zongker et al. 1999] observe how known light patterns refract through and reflect from objects, which imposes constraints on the possible surface curvature. [Vasilyev et al. 2011] showed that detailed surface geometry could be reconstructed when the camera was fixed to the surface, and both moved through a known scene, inducing a specular-flow field. Our approach is partly similar, inverting the capture process so the objects stay stationary, but the hand-held camera moves around.

B1.2.2 Image-based Rendering (IBR)

IBR systems have their roots in the early work on view interpolation [Chen and Williams 1993], where the goal was to create compelling transitions between pairs of images. This idea quickly developed towards computing completely new views from a set of input images. Debevec *et al.* [1996] show that combining a coarse geometric model with texture mapping that is specifically computed for the current view can create realistic renderings of simple, architectural scenes from viewpoints away from the original photographs. A general theory was presented in the Light Field and Lumigraph papers [Levoy and Hanrahan 1996; Gortler et al. 1996]. These frameworks interpolate rays sampled from a four-dimensional light field to generate the required rays (and thus pixels) for a novel view. Using 3D proxies for a given scene improves the quality of this basic technique. These proxies provide a more accurate description of the motion of corresponding pixels and hence a higher quality interpolation is possible [Gortler et al. 1996]. Many other techniques build on this idea and present improved systems.

Narayanan *et al.* [1998] capture images using a camera dome and compute global and per-view geometry models for rendering. Given a long video sequence of a scene, Heigl *et al.* [1999] hierarchically compute a piece-wise planar scene reconstruction from triples of camera views and use an image-based approach to render novel views. In a similar vein, [Buehler et al. 2000] present the unstructured Lumigraph, which allows users to move away from sparse and unorganized input photographs. Lhuillier and Quan [2003] first reconstruct per-view depth maps and introduce a consistent triangulation of depth maps for pairs of views. Evers-Senne and Koch [2003] reconstruct dense but incomplete depth maps from a dense set of images captured under controlled conditions. At render time, they project a subset of these depth maps into the novel view and smoothly interpolate any remaining holes. Hornung and Kobbelt [2009] improve on this by using feature-aware particles and performing pixel-accurate color accumulation in a fully GPU-based pipeline. All these techniques rely on relatively complete geometric representations and cannot handle viewpoint changes that are far away from the input data. In these cases the missing depth information causes holes in the geometry and significant visual artifacts.

Recent work by consortium members has provided significant improvements to previous techniques. For example, the work of **Goesele *et al.* [2010]** aims to remove such holes by explicitly rendering the uncertainties of the geometry. They are, however, again restricted to transitions between two views. The work of **Chaurasia et al. [2011, 2013]**, proposes methods which allows users to move far from the input views. This is achieved by using shape-preserving warps using occlusion boundaries, based on an initial oversegmentation **Chaurasia et al. [2013].**

The second challenge is to allow modification of the lighting of the captured images. Evidently, most image-based rendering methods are restricted to the lighting at the time the photos were taken: our goal is to allow modification of the lighting conditions so that a single capture can be used to reproduce multiple lighting conditions of the same scene or scene objects. This is a hard problem, with no easy-to-use solution available to this date. Debevec [1998] was the first to describe a capture and rendering setup that reconstructs the lighting situation of the scene in order to insert synthetic objects without noticeable visual artifacts. In further work several acquisition setups (light stages) have been proposed to estimate the complete reflectance of objects [Debevec 2000]. These systems usually capture a set of viewpoints under a dense sampling of incident lighting directions. This is, however, only possible in a controlled and often very expensive laboratory environment, where several lighting scenarios can be synthesized. For CR-PLAY we need to capture under uncontrolled illumination settings.

Another possibility to achieve relightable models is based on *intrinsic image* decompositions of the input photographs, i.e., into illumination and reflection layers. Intrinsic images can be computed from a single photograph [Horn 1986] [Tappen 2005], which is, however, a very underconstrained problem since many combinations of illumination and reflection layers can produce the same image. Solving this problem thus requires additional assumptions about the scene.
Richer and more robust decompositions can be achieved by using a set of images. Sunkavalli et al. [2007] use timelapse sequences to estimate a shadow mask and separate sky and sun illumination. **Haber et al. [2009]** estimate BRDFs and distant illumination in 3D scenes reconstructed with multi-view stereo. [**Ackermann et al. 2012]** reconstruct reflectance and surface orientation from web cam data. Other approaches, developed at

participating institutes, also provide excellent results for multi-view input data, captured with a specific setup and under a single illumination **[Laffont et. al. 2013]**, or even for completely unstructured photo collections **[Laffont et al. 2012]**.

In some settings, such as with tall buildings or dense forests, only nearby samples can be captured carefully, yet IBR must be performed for both near and distant surfaces. Here, the history of image-texture synthesis, as summarized in [Wei et al. 2009], reveals that non-parametric synthesis such as [Efros and Freeman 2001], while quite naive, easily overtakes more sophisticated parametric methods both in terms of quality and generalizability to various stochastic textures. [Kopf et al. 2007] show that, with some modification, a similar paradigm applies also to volume-synthesis. On the other hand, bidirectional Texture Functions can be combined or “decorated” [Zhou et al. 2005] and from our own efforts, even edited **[Kautz et al. 2007]**, but are difficult for users to acquire. Existing tools for 3D content editing and creation, such as modeling and sculpting packages, are prohibitively difficult and non-intuitive for typical users, who often already have examples of the thing they wish to model, or at least arm’s-length access to some part. One solution, albeit incredibly labor-intensive, is to specify material simulation models and parameters, such as weathering-due-to-rain [Lu et al. 2007] on explicitly modeled volumes or sketch-based models [Olsen et al. 2009]

B1.2.3 Video-based Rendering (VBR)

Video Browsing Video-based rendering methods synthesize new views from videos of a scene. The pioneering work of Andrew Lippman [1980] realized one of the first systems for interactive navigation through a database of images. Subsequent research attempted to automate this process. For instance, Kimber et al.'s *FlyAbout* [2001] captured panoramic videos by moving a 360° camera along continuous paths and synthesized novel views by mosaicing. Users chose a path through a constrained set of automatically pre-computed branching points, and at these points only novel view synthesis is required. We describe heuristics, investigated through a user study, to select appropriate transition rendering styles. In a *telepresence* context, McCurdy and Griswold's *Realityflythrough* [2005] establishes connections between videos from mobile devices based on GPS information and provides a simple transition between overlapping videos in a manner similar to [Snavely et al. 2006]. At transitions, videos are projected onto their respective image planes. Further related approaches exist for navigating through real scenes captured in videos, such as the work by Saurer et al. [2010] or our work on Videoscapes **[Tompkin et al. 2012]**. However, these methods rely on a constrained capture environment (e.g., special hardware or confined spatial locations), which facilitates processing and rendering. All these approaches take video as input and synthesize (novel) views from the input data. In contrast to these approaches, we seek to recreate dynamic environments from videos, as opposed to touring through captured videos.

Dynamic VBR Early work focused on model-based representations for stochastic temporal textures [Szummer and Picard 1996; Bar-Joseph et al. 2001; Soatto et al. 2001], which could synthesize dynamic sequences of phenomena such as smoke and water. Video textures [Schödl et al. 2000] was the first approach to create dynamic textures of arbitrary objects under repetitive and stochastic motions. Frames from a training video are reordered and repeated indefinitely such that a new generated video is never exactly the same as the input.

[Schödl and Essa 2002] also provide an extension to generate video textures of moving objects viewed from static cameras. [Kwatra et al. 2003] create better transitions within video textures by finding minimum cost seams through a window around similar frames, and so enable video textures for more difficult surfaces such as water. [Agarwala et al. 2005] extend video textures to panoramic imagery by rotating a single camera from a single position and solving for y, t video volume slices. [Bhat et al. 2004] demonstrated that video of certain repetitive events, especially video of fluid flowing, could be edited to achieve special effects like widening a waterfall or stretching out a water-spout, but it is limited to fixed cameras.

Free-viewpoint video [Zitnick et al. 2004; Magnor 2005] techniques operate in a similar context as this proposal. A scene is usually captured with multiple video cameras. The videos are analyzed and 3D geometry is reconstructed. Novel viewpoints are rendered from the videos and the (per-frame) 3D geometry. Some methods only rely on dense per-pixel image correspondences [Lipski et al. 2010] instead of full 3D geometry. However, in contrast to these free-viewpoint video methods, we focus on dynamic stochastic objects, such as trees, which are difficult to reconstruct.

Sand et al. [Sand and Teller 2004] describe a method for bringing two videos, captured from similar spatial but different temporal locations, into spatiotemporal alignment. The method relies on regularized sparse correspondence formed from static image regions, and it explicitly rejects correspondence on dynamic objects. However, we specifically require correspondence on dynamic objects. Applying this technique to our problem will not find correspondence for motion in dynamic scenes, and will cause ghosting. [Knorr et al. 2008] turn monocular video into super-resolved stereo and multi-view video. However, as this approach is dependent on structure-from-motion, it is not suitable for finding correspondence on dynamic objects.

Uyttendaele et al. [Uyttendaele et al. 2004] present an interactive system to navigate omnidirectional video of real-world environments. They perform object replacement to enhance the static world viewed when the camera position is stopped, e.g., replacing a static fireplace with a video texture – a limited approach, as it assumes planar dynamic objects. The technique proposed by **[Ballan et al. 2010]** enables blending between different videos showing a single spatially confined scene or event. They assume a scene model with a billboard in the foreground and 3D geometry in the background. The background is reconstructed from additional community photos of the scene, and the video cameras are calibrated w.r.t. the background model.

The system is state of the art, but is tailored to spatially confined sets of videos that all see the same event at the same time from converging camera angles.

Our work on interactive viewpoint video textures **[Levieux et al. 2012]** takes an initial step towards rendering video-based dynamic objects. It essentially uses video textures in space and time, but requires careful sequential capture and cannot extract dynamic objects.

References

ACKERMANN, J., LANGGUTH, F., FUHRMANN, S., GOESELE, M., 2012. Photometric stereo for outdoor webcams. In: Proceedings of CVPR 2012, Providence, USA.

AGARWALA, A., ZHENG, K. C., PAL, C., AGRAWALA, M., COHEN, M., CURLESS, B., SALESIN, D., AND SZELISKI, R. 2005. Panoramic video textures. *ACM Trans. Graph. 24,* 3 (July), 821–827.

ALIAGA, D., FUNKHOUSER, T., YANOVSKY, D., AND CARLBOM, I. 2003. Sea of images. *IEEE Computer Graphics and Applications 23*, 6, 22–30.

ALLDRIN, N. G., ZICKLER, T., AND KRIEGMAN, D. 2008. Photometric stereo with non-parametric and spatially-varying reflectance. In *CVPR*.

BALLAN, L., BROSTOW, G., PUWEIN, J., AND POLLEFEYS, M. 2010. Unstructured video-based rendering: Interactive exploration of casually captured videos. *ACM Trans. Graph. (Proc. SIGGRAPH) 29*, 3, 87:1–87:11.

BAR-JOSEPH, Z., EL-YANIV, R., LISCHINSKI, D., AND WERMAN, M. 2001. Texture mixing and texture movie synthesis using statistical learning. vol. 7. IEEE Computer Society, 120–135.

BHAT, K. S., SEITZ, S. M., HODGINS, J. K., AND KHOSLA, P. K. 2004. Flow-based video synthesis and editing. SIGGRAPH ’04, 360–363.

BOUGUET, J.-Y., AND PERONA, P. 1998. 3d photography on your desk. ICCV, 43–50.

BROSTOW, G. J., HERNÁNDEZ, C., VOGIATZIS, G., STENGER, B., AND CIPOLLA, R. 2011. Video normals from colored lights. *IEEE Trans. Pattern Anal. Mach. Intell. 33*, 10, 2104–2114.

BUEHLER, C., BOSSE, M., MCMILLAN, L., GORTLER, S., AND COHEN, M. 2001. Unstructured lumigraph rendering. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, SIGGRAPH ’01, 425–432.

BURMEISTER, O.K. Usability Testing: Revisiting Informed Consent procedures for testing Internet sites, 2nd Australian Institute of Computer Ethics Conference, AICE2000, Camberra, December 2000, ACE, 2001.

CHAURASIA, G., SORKINE, O., AND DRETTAKIS, G. 2011. Silhouette-aware warping for image-based rendering. In *Proceedings of the Twenty-second Eurographics conference on Rendering*, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, EGSR’11, 1223–1232.

CHAURASIA, G., DUCHENE S., SORKINE, O., AND DRETTAKIS, G. ., Depth Synthesis and Local Warps for Plausible Image-based Navigation. (accepted with minor revisions, ACM Transactions on Graphics), 2013.

CHEN, S. E., AND WILLIAMS, L. 1993. View interpolation for image synthesis. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, SIGGRAPH ’93, 279–288.

DAVIS, A., LEVOY, M., AND DURAND, F. 2012. Unstructured light fields. *Comput. Graph. Forum 31*, 2, 305–314.

DEBEVEC, P. E., TAYLOR, C. J., AND MALIK, J. 1996. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, SIGGRAPH ’96, 11–20.

DEBEVEC, P., HAWKINS, T., TCHOU, C., DUIKER, H.-P., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, SIGGRAPH '00, 145–156.

DEBEVEC, P. 1998. Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, SIGGRAPH '98, 189–198.

DUMAS, J. S. AND REDISH, J. C. 1994: A practical guide to usability testing, Norwood: Ablex Publishing Corp.

EBERT, D.S., 2003. *Texturing & Modeling: A Procedural Approach*, Morgan Kaufmann.

EFROS, A. A., AND FREEMAN, W. T. 2001. Image quilting for texture synthesis and transfer. In *SIGGRAPH*, 341–346.

EVERS-SENNE, J.-F., AND KOCH, R. 2003. Interactive rendering with view-dependent geometry and texture. In *ACM SIGGRAPH 2003 Sketches & Applications*, ACM, New York, NY, USA, SIGGRAPH '03, 1–1.

GOESELE, M., ACKERMANN, J., FUHRMANN, S., HAUBOLD, C., KLOWSKY, R., STEEDLY, D., AND SZELISKI, R. 2010. Ambient point clouds for view interpolation. In *ACM SIGGRAPH 2010 papers*, ACM, New York, NY, USA, SIGGRAPH '10, 95:1–95:6.

GOESELE, M., SNAVELY, N., CURLESS, B., HOPPE, H., SEITZ, S. 2007. Multi-View Stereo for Community Photo Collections. In: Proceedings of ICCV 2007, Rio de Janeiro, Brasil.

GORTLER, S. J., GRZESZCZUK, R., SZELISKI, R., AND COHEN, M. F. 1996. The lumigraph. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, SIGGRAPH '96, 43–54.

HABER, T., FUCHS, C., BEKAERT, P., SEIDEL, H.-P., GOESELE, M., AND LENSCH, H. P. A. 2009. Relighting objects from image collections. In *IEEE CVPR*, IEEE, 627–634.

HASTINGS, E.J., GUHA, R.K. AND STANLEY, K.O., 2009. Automatic Content Generation in the Galactic Arms Race Video Game. *IEEE Transactions on Computational Intelligence and AI in Games*, 1(4), pp.245–263.

HEIGL, B., KOCH, R., POLLEFEYS, M., DENZLER, J., ANDGOOL, L. J. V. 1999. Plenoptic modeling and rendering from image sequences taken by hand-held camera. In *Mustererkennung 1999*, 21. DAGM-Symposium, Springer-Verlag, London, UK, UK, 94–101.

HERNÁNDEZ, C., VOGIATZIS, G., AND CIPOLLA, R. 2008. Multiview photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell. 30*, 3, 548–554.

HERTZMANN, A., AND SEITZ, S. M. 2005. Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *IEEE Trans. Pattern Anal. Mach. Intell. 27*, 8 (Aug.), 1254–1264.

HIGO, T., MATSUSHITA, Y., JOSHI, N., AND IKEUCHI, K. 2009.A hand-held photometric stereo camera for 3-d modeling. In *ICCV*.

HORN, B. K. 1986. *Robot Vision*, 1st ed. McGraw-Hill Higher Education.

HORNUNG, A., AND KOBBELT, L. 2009. Interactive pixel-accurate free viewpoint rendering from images with silhouette aware sampling. *Computer Graphics Forum* 28, 8, 2090–2103.

JOSHI, N., AND KRIEGMAN, D. 2007. Shape from varying illumination and viewpoint. In *ICCV*.

KAUTZ, J., BOULOS, S., AND DURAND, F. 2007. Interactive editing and modeling of bidirectional texture functions. In *ACM SIGGRAPH 2007 papers*. ACM, New York, NY, USA, SIGGRAPH '07.

KIMBER, D., FOOTE, J., AND LERTSITHICHAI, S. 2001. Flyabout: spatially indexed panoramic video. In *Proc. ACM Multimedia*, 339–347.

KNORR, S., KUNTER, M., AND SIKORA, T. 2008. Stereoscopic 3d from 2d video with super-resolution capability. *Image Commun. 23*, 9 (Oct.), 665–676.

KOPF, J., FU, C.-W., COHEN-OR, D., DEUSSEN, O., LISCHINSKI, D., AND WONG, T.-T. 2007. Solid texture synthesis from 2d exemplars. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2007) 26*, 3, 2:1–2:9.

KWATRA, V., SCH¨ ODL, A., ESSA, I., TURK, G., ANDBOBICK, A. 2003. Graphcut textures: Image and video synthesis using graph cuts. In *ACM SIGGRAPH 2003 Papers*, SIGGRAPH '03, 277–286.

LAFFONT, P.-Y., BOUSSEAU, A., PARIS, S., DURAND, F., AND DRETTAKIS, G. 2012. Coherent intrinsic images from photo collections. *ACM Trans. Graph. 31*, 6 (Nov.), 202:1–202:11.

LAFFONT, P.-Y., BOUSSEAU, A., AND DRETTAKIS, G. 2013. Rich intrinsic image decomposition of outdoor scenes from multiple views. *IEEE Transactions on Visualization and Computer Graphics 19*, 2 (Feb.), 210–224.

LEVIEUX, P., TOMPKIN, J., AND KAUTZ, J. 2012. Interactive viewpoint video textures. In *Conference on Visual Media Production (CVMP)*, 11–17.

LEVOY, M., AND HANRAHAN, P. 1996. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, SIGGRAPH '96, 31–42.

LHUILLIER, M., AND QUAN, L. 2003. Image-based rendering by joint view triangulation. *IEEE Trans. Cir. and Sys. for Video Technol. 13*, 11 (Nov.), 1051–1063.

LIPPMAN, A. 1980. Movie-maps: An application of the optical videodisc to computer graphics. *Computer Graphics (Proc. SIGGRAPH) 14*, 3, 32–42.

LIPSKI, C., LINZ, C., BERGER, K., SELLENT, A., AND MAGNOR, M. 2010. Virtual video camera: Image-based viewpoint navigation through space and time. *Computer Graphics Forum 29*, 8 (Dec.), 2555–2568.

LU, J., GEORGHIADES, A. S., GLASER, A., WU, H., WEI, L.-Y., GUO, B., DORSEY, J., AND RUSHMEIER, H. 2007. Context-aware textures. *ACM Trans. Graph. 26*, 1 (Jan.).

MAC AODHA, O., CAMPBELL, N. D., NAIR, A., AND BROSTOW, G. J. 2012. Patch based synthesis for single depth image super-resolution. In *ECCV (3)*, 71–84.

MAGNOR, M. 2005. *Video-based Rendering*. A K Peters.

MCCURDY, N. J., AND GRISWOLD, W. G. 2005. A systems architecture for ubiquitous video. In *Proc. International Conference on Mobile Systems, Applications, and Services*, 1–14.

NARAYANAN, P. J., RANDER, P. W., AND KANADE, T. 1998. Constructing virtual worlds using dense stereo. In *Proceedings of the Sixth International Conference on Computer Vision*, IEEE Computer Society, Washington, DC, USA, ICCV '98, 3–.

NEHAB, D., RUSINKIEWICZ, S., DAVIS, J., AND RAMAMOORTHI, R. 2005. Efficiently combining positions and normals for precise 3D geometry. *ACM Transactions on Graphics (Proc. Of ACM SIGGRAPH 2005) 24*, 3 (Aug.).

NEWCOMBE, R. A., IZADI, S., HILLIGES, O., MOLYNEAUX, D., KIM, D., DAVISON, A. J., KOHLI, P., SHOTTON, J., HODGES, S.,AND FITZGIBBON, A. 2011. Kinectfusion: Real-time dense surface mapping and tracking. ISMAR '11, 127–136.

OLSEN, L., SAMAVATI, F. F., SOUSA, M. C., AND JORGE, J. A. 2009. Technical section: Sketch-based modeling: A survey. *Computers & Graphics 33*, 1, 85–103.

PERLIN, K., 1985. An image synthesizer. In *ACM SIGGRAPH Computer Graphics*. pp. 287–296.

PONGNUMKUL, S., WANG, J., AND COHEN, M. 2008. Creating map-based storyboards for browsing tour videos. In *Proc. ACM Symposium on User Interface Software and Technology*, 13–22.

REYNOLDS, M., DOBOŠ, J., PEEL, L., WEYRICH, T., AND BROSTOW, G. J. 2011. Capturing time-of-flight data with confidence. In *CVPR*.

SAND, P., AND TELLER, S. 2004. Video matching. In *ACM SIGGRAPH 2004 Papers*, ACM, New York, NY, USA, SIGGRAPH ’04, 592–599.

SAURER, O., FRAUNDORFER, F., AND POLLEFEYS, M. 2010. OmniTour: Semi-automatic generation of interactive virtual tours from omnidirectional video. In *Proc*. 3DPVT, 1–8.

SCHÖDL, A., ANDESSA, I. A. 2002. Controlled animation of video sprites. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA ’02, 121–127.

SCHÖDL, A., SZELISKI, R., SALESIN, D. H., AND ESSA, I. 2000. Video textures. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH ’00, 489–498.

SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2006. Photo tourism: exploring photo collections in 3D. *ACM Trans. Graph. (Proc.SIGGRAPH) 25*, 3, 835–846.

SOATTO, S., DORETTO, G., AND WU, Y. 2001. Dynamic textures. In *Proceedings of the International Conference on Computer Vision*, vol. 2, 439–446.

SUNKAVALLI, K., MATUSIK, W., PFISTER, H., AND RUSINKIEWICZ, S. 2007. Factored time-lapse video. In *ACM SIGGRAPH 2007 papers*, ACM, New York, NY, USA, SIGGRAPH ’07.

SZUMMER, M., AND PICARD, R. W. 1996. Temporal texture modeling. In *IEEE Intl. Conf. Image Processing*, vol. 3, 823–826.

TAPPEN, M. F., FREEMAN, W. T., AND ADELSON, E. H. 2005. Recovering intrinsic images from a single image. *IEEE Trans. Pattern Anal. Mach. Intell. 27,* 9 (Sept.), 1459–1472.

THUE, D., BULITKO, V. AND SPETCH, M., 2008. PaSSAGE: A Demonstration of Player Modelling in Interactive Storytelling. In *Proceedings of the Fourth Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE ’08)*. Palo Alto, CA: AAAI Press, pp. 227–228.

TOMPKIN, J., KIM, K. I., KAUTZ, J., AND THEOBALT, C. 2012. Videoscapes: Exploring sparse, unstructured video collections. *ACM Trans. Graphics (Proc. SIGGRAPH) 31*, 4, 68:1–68:12.

UMMENHOFER, B., AND BROX, T. 2012. Dense 3D reconstruction with a hand-held camera. In *Pattern Recognition (Proc.DAGM).*

UYTTENDAELE, M., CRIMINISI, A., KANG, S. B., WINDER, S., SZELISKI, R., AND HARTLEY, R. 2004. Image-based interactive exploration of real-world environments. *IEEE Computer Graphics and Applications 24,* 3, 52–63.

VASILYEV, Y., ZICKLER, T., GORTLER, S., AND BEN-SHAHAR, O. 2011. Shape from specular flow: Is one flow enough? CVPR, 2561–2568.

WEI, L.-Y., LEFEBVRE, S., KWATRA, V., AND TURK, G. 2009. State of the art in example-based texture synthesis. In *Eurographics 2009, State of the Art Report, EG-STAR.*

WOODHAM, R. 1980. Photometric method for determining surface orientation from multiple images. In *Optical Eng.*, vol. 19, 139–144.

ZHANG, L., CURLESS, B., HERTZMANN, A., AND SEITZ, S. M. 2003. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. In *ICCV*, 618–625.

ZHOU, K., DU, P., WANG, L., MATSUSHITA, Y., SHI, J., GUO, B.,AND SHUM, H.-Y. 2005. Decorating surfaces with bidirectional texture functions. *IEEE Transactions on Visualization and Computer Graphics 11*, 5.

ZITNICK, C. L., KANG, S. B., UYTTENDAELE, M., WINDER, S., AND SZELISKI, R. 2004. High-quality video view interpolation using a layered representation. *ACM Trans. Graph. 23,* 3, 600–608.

ZONGKER, D. E., WERNER, D. M., CURLESS, B., AND SALESIN, D. H. 1999. Environment matting and compositing. In *Proceedings of ACM SIGGRAPH 99*, 205–214.