Using sparse optical flow for multiple Kinect applications

Kai Berger
OeRC
University of Oxford
Oxford, United Kingdom
Email: kai.berger@oerc.ox.ac.uk

Yannic Schroeder, Stefan Guthe
TU Braunschweig
Braunschweig, Germany
Email: {schroeder,guthe}@cg.cs.tu-bs.de

Marc Aurel Kastner
TU Braunschweig
Braunschweig, Germany
Email: marc.kastner@gmail.com

***Abstract*—The use of Multiple Microsoft Kinects has become prominent in the last two years and enjoyed widespread acceptance. While several work has been published to mitigate quality degradations in the precomputed depth image, this work focuses on employing an optical flow suitable for dot patterns as employed in the Kinect to retrieve subtle scene data alterations for reconstruction. The method is employed in a multiple Kinect vision architecture to detect the interface of propane flow around occluding objects in air.**

I. Introduction

With the advent of the Microsoft Kinect not only the human-computer interaction in the consumer market has been shifted to a new level, the interaction based on motion capturing, but also computer vision based research has experienced a sparked since the release. Ranging from advances in motion capturing and real-time self localisation and mapping to robust face capturing, the bandwidth of the research volume related to the Kinect is surprisingly broad. While most approaches have focused on preprocessed data of the sensor, i.e. a considerably accurate depth estimate of the scene, only few have actually considered the raw IR stream for their data processing. This article exploits the use of the IR footage, containing greylevel information about the scene together with sparsely distributed spots emitted from the IR laser component of the Kinect. While the nondisclosed algorithm of the Kinect computes a dense depth image of the scene from considerably shifted positions of the laser spots, the proposed algorithm will detect subtle alterations introduced on the captured spots image, e.g. in the scenario of small light path deviations introduced by a refractive medium. The resulting information is equivalent to an optical flow image as it quantitatively denotes the alteration, i.e. pixel movement, introduced on each spot. It, however, differs from state-of-the-art optical flow images as it considers only a sparse discrete set of spots, projected into the scene from a IR laser component and captured with the IR camera, which are represented by its barycentric position in the image and its distincitve shape. Furthermore the information about the projected pattern allow a computation *independent* from the actual resolution of the underlying image stream. We will test the algorithm in a capturing setup consisting of multiple Microsoft Kinects. The refractive index gradient in the scene is generated by letting propane flow through the intersecting volume of their viewing cones, Fig. I.

The remainder of the article is structured as follows: After a revision of the state-of-the-art in active-light-based computer vision research with a focus on optical flow capturing, Sect. II, we examine the characteristics of the spot pattern projected by the Kinect and derive an abstract model in Sect. III that allows us to compute an accurate estimate of the spot position on the image plane *independent* from the actual resolution of the underlying image stream. Afterwards, we introduce the subpixel accurate optical flow computation for detecting subtle changes in the captured IR stream, Sect. IV. Note, that the proposed algorithms differs significantly from state-of-the-art optical flow computation as it considers only sparse samples of the input images. A global optimization based approach is therefore not feasible.

The presented method is then applied in the scenario of capturing the interface of two-phase gas flows, Fig. I, Sect. V and Sect. V-B. Finally we discuss the results, Sect. VI, and conclude, Sect. VII.

II. Related work

A variety of approaches have been proposed to capture and visualize the dynamics of gases and fluids, e.g. in windtunnels. Mostly, two-phase flows are hard to discriminate by imaging tools, because their color appearance is quite similar. A straight-forward approach is to enhance the contrast by inserting seeding particles into the flow. Their buoyancy and density has to match the measured fluids. The particles are usually illuminated in a way that only a 2d intersection plane is imaged. The technique is known as Particle Image Velocimetry (PIV), an introduction can be found in the monograph by [21]. Perpendicular flows can be detected stereoscopically [9]. Another approach, the Particle Tracking Velocimetry (PTV) introduced by Cowen and Monismith [13] is similar to the PIV but incorporates tracking algorithms for particles over several images. Bendick et al. [12] rened colored particles to make the distinction easier. A usual drawback of that approach is, that the particles deviate slightly from the flow by inertial

Fig. 1. We examine small pixel deviations in the IR-stream of the Microsoft Kinect as they are introduced by an index gradient present in the scene. In our setup we place a propane gas nozzle (left) in the center of three Kinects (middle) and place projection walls at a fixed distance opposite to each Kinect. The index gradient causes light path deviations both between the projector and the projection wall and between the projection wall and the camera (indicated by coloured cones). To reconstruct the underlying gas flow using the proposed optical flow algorithm we register each sensor to one common world space (right) and reconstruct the visual from silhouettes enclosing the deflected pixel regions in each sensor's IR-image.

force. In Laser Doppler Anemometry (LDA) a coherent laser beam is split into two beams focused on the same point in a flowing volume. A particle passing this point reflects the lasers light and a photodetector measures the resulting interference pattern [14]. Other approaches are non-invasive. Atcheson et al. [5] designed a high-frequency background pattern [6] for Background-Oriented-Schlieren (BOS) detection of gas flows. This approach was refined by Berger et al. [8] to enable the capturing of gas flows with occluding objects. They however showed, that purely image-based methods are limited, and occluders could not be segmented out crisply. Therefore, Berger et al. [10] introduced the usage of Kinects for capturing gas flows around occluders. They imaged gas introduced dierences in the captured depth stream. When the scene was captured with multiple Kinects a reconstruction of the visual hull of the gas was possible, but it was limited to superpixel resolution. Their work is the closest to ours, as we consider the same capturing scenario. We however advance it in the following way: We introduce a new sparse spot-based optical flow to work directly on the IR input stream of the sensor and thus are to work on subpixel accurate light-path deviation data. Furthermore, we do not simply reconstruct the visual hull of the gas [10], but seek to geometrically reconstruct the interface between the flowing gas and the surrounding air by incorporating a total variation approach.

III. Spot model derivation

As the pattern projected by the Microsoft Kinect has a unique distinctive shape, it allows for analysing the geometric properties of the underlying pattern elements, the *spots*. Each projected spot pattern is set up in the same way: in a 3×3 repeated regular matrix certain circularly shaped holes in a mask are filled with opaque matter, while others are left unfilled, so that light beams may pass through. This matrix is applied to the optical element of the Kinect's IR emitter such that a coded light pattern emerges and projects into the scene. We analyse the IR-pattern by letting the emitter projected orthogonally onto a wall at $3m$ distance and capturing a statistically significant area of the wall with a digatal single lens reflex (dSLR) camera at a resolution of $4.368 \times 2.912px$ at a distance of $2m$. This way, one imaged projected spot would approximately comprise $128 \times 128px$. After image rectification we calculate the mean intensity distribution over all captured spots and the moments with the intention to fit a suitable continuous 2d distribution function that characterises the properties of all spots best. The function would then represent the mean spot. This is reasoned by the fact that we want to perform the optical flow computation on the image plane independent from the actual image resolution. Note, that this mean distribution would be used for fitting to each spot in each IR-image in order to locate the center position of each spot. The computed discrete distribution resembles a 2d Gaussian profile. However, the distribution shows a skew behavior and 8 small local peaks in the neighborhood of the mean. The latter is due to the factory process of the matrix, the holes filled with opaque matter may transmit a small but measurable intensity of light. We opt for representing the continuous distribution by a weighted combination of continuous 2d skew Gaussian distributions. Such multivariate distribution can be formalized as follows [23]:

Consider a multivariate variable $\mathcal{Z}$ with each component being skew-normal with skewness vector $\lambda, \lambda_i \in (-\infty, \infty)$, written as:

$$\mathcal{Z} \sim \mathcal{SN}_k(\lambda, \Phi) \tag{1}$$

with multivariate density function f_k of dimensionality k

$$f_k(z) = 2\phi_k(z; \Omega)\Phi(\alpha^T z) \;\; (z \in R^k), \tag{2}$$

where ϕ denotes the $\mathcal{N}(O, 1)$ density function and Φ denotes the $\mathcal{N}(O, 1)$ distribution . Ω is defined as

$$\Omega = diag((1-\delta_1^2)^{\frac{1}{2}}, ..., (1-\delta_k^2)^{\frac{1}{2}})(\Psi+\lambda\lambda^T)diag((1-\delta_1^2)^{\frac{1}{2}}, ..., (1-\delta_k^2)^{\frac{1}{2}}) \tag{3}$$

with Ψ the $k \times k$ correlation matrix and

$$\delta = \frac{\lambda}{(1+\lambda^2)^{\frac{1}{2}}} \tag{4}$$

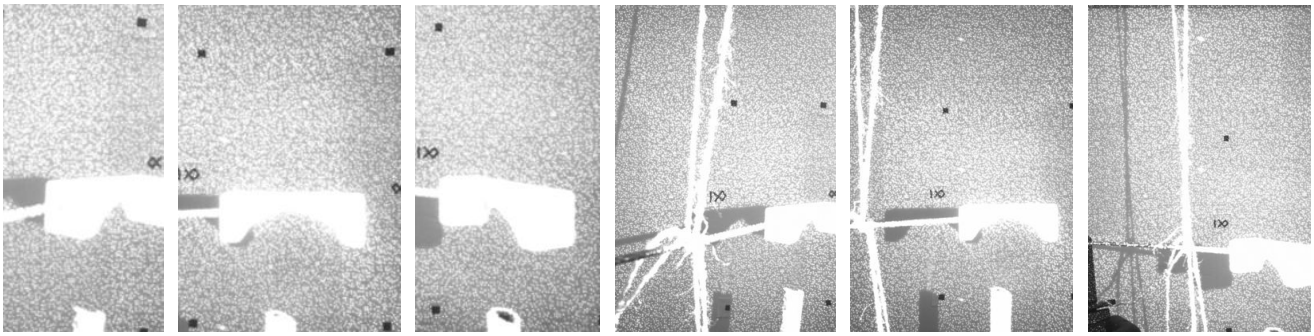

Fig. 2. The input images of the capturing setup for two-phase gas flows (left three) and their warped projector images (right three). The gas flows out of a valve (bottom) and is imaged from three Kinects placed in an half-arc around it. An occluder is placed above the valve to introduce turbulence to the gas flow. We examine the both the deflections introduced on the way between projector and wall (right three shadowgraphs) and on the way between wall and camera (left three).

α^T is defined as

$$\alpha^T = \frac{\lambda^T \Psi^{-1} diag((1-\delta_1^2)^{\frac{1}{2}}, ..., (1-\delta_k^2)^{\frac{1}{2}})^{-1}}{(1+\lambda^T \Psi^{-1} \lambda)^{\frac{1}{2}}} \quad (5)$$

We will focus on the bivariate case in order to model the intensity distribution on the image plane. The density function can be expressed as

$$f_2(z_1, z_2) = 2\psi_2(z_1, z_2; \omega)\Psi(\alpha_1 z_1 + \alpha_2 z_2), \quad (6)$$

with ω denoting the off-diagonal element of Ω. α_1 and α_2 can be expressed as:

$$\alpha_1 = \frac{\delta_1 - \delta_2 \omega}{((1-\omega^2)(1-\omega^2-\delta_1^2-\delta_2^2+2\delta_1\delta_2\omega))^{\frac{1}{2}}} \quad (7)$$

and

$$\alpha_2 = \frac{\delta_2 - \delta_1 \omega}{((1-\omega^2)(1-\omega^2-\delta_1^2-\delta_2^2+2\delta_1\delta_2\omega))^{\frac{1}{2}}} \quad (8)$$

Note, that ω satisfies

$$\delta_1\delta_2 - ((1-\delta_1^2)(1-\delta_2^2))^{\frac{1}{2}} < \omega < \delta_1\delta_2 + ((1-\delta_1^2)(1-\delta_2^2))^{\frac{1}{2}} \quad (9)$$

IV. Proposed flow computation for captured spot-pattern based active light video data

Based on the distribution, that we derived in Sect. III we compute frame-to-frame alterations to emitted spot-pattern introduced by the scene content. Note, that this sparse flow computation has a focus on subtle movements that may be smaller than one pixel of the captured image resolution and thus distinguishes from state-of the art approaches that focus on the local disparity estimate in order to compute scene depth. We discretize Eqn. 6 in a $128 \times 128px$ kernel that we then use to characterise the mean spot, Fig. 3. We filter the input images to determine the center position for each spot by fitting the discretized kernel to all intensity peaks present in the input image. We look for the intensity distribution of the dot pattern only once to build our $128 \times 128px$ kernel of a generic spot. This kernel is then used at each time a Kinect IR stream is read out later.

The sparse optical flow is then computed between the computed spot center positions for two subsequent images in the IR-stream in a least squares sense, assuming that the pixel deviations introduced on the spot pattern are smaller than the radial distance between two spots in the image. The local optimum is found by applying a gradient descent approach

$$x_{n+1} = x_n - \gamma_n \nabla F(x_n), n \geq 0, \quad (10)$$

where x_n spans the domain, in this case the position on the image plane, and F is a bivariate function with defined and differentiable neighborhood. The gray level image is denoted F, and the gradient can be provided as a 2D vector. To account for the low sampling rate, the region of interest, usually a $5 \times 5px$ region, is upsampled. The stepsize γ is set to a small value, usually $.5px$. The convergence criterion is an intensity difference of < 0.01 and the maximum number of iterations is set to 100. We assume small spot deviations in the rang of < 1 to 3 pixel in x and y direction. Note, again, that the resulting optical flow is sparse as well.

V. Application: Capturing of two-phase gas flows

In our setup we place three Kinects in an half-arc around the gas flow with projection walls placed opposite to each Kinect at $\approx 2m$ distance. We may introduce occluders into the flow. The Kinects are Calibrated to a common world-space and aligned to frame-accuracy. We let the gas valve vent for several seconds and capture the gas flow with the Kinects. The captured images of the IR stream, Fig. 2, are processed in the following.

A. Projector Image reprojection

We exploit both the sensor information of the infrared camera and the infrared emitter for the reconstruction. This way we can also use the shadowgraph regions in the IR-image for optical flow calculation and consequently for a silhouette generation that helps refining the final visual hull enclosing the gas flow. Usually the detected flow directions in the shadowgraph regions are inverted compared

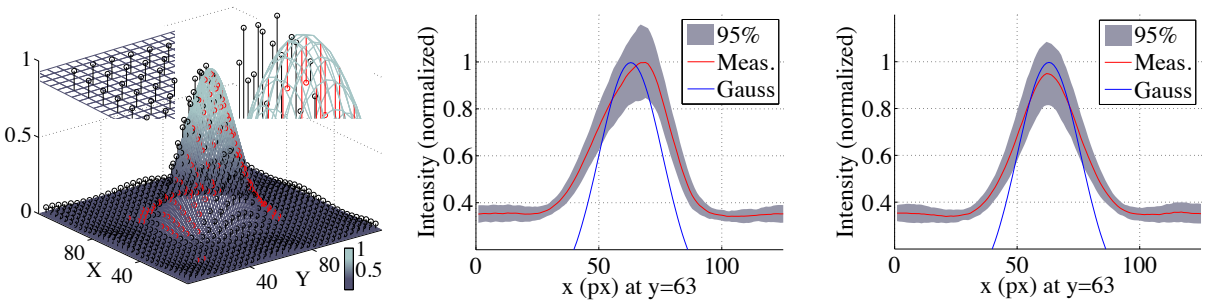

Fig. 3. The intensity profile a measured spot in the projected IR-pattern differs considerably from the profile of a standard Gaussian laser beam (left, black and red error bars and slices, middle and right). In this plot a 50mW laser spot is assumed, that is shone onto a planar surface from 3m distance and sampled with a 125 125px grid. We thus discretize a skew Gaussian distribution to a 125 125px kernel and fit it to the spots in the captured spot pattern to arrive at a discretized mean intensity distribution. The discretized kernel is later used for our spot-based optical computation.

to the regions that image light path deviations between the projection wall and the camera. The projector image is generated by reprojecting the infrared camera image into the projector image space. The calibration between projector and camera is done by identifying certain spot sub-patterns of the projected pattern in the recorded image that form a rectangle in the projector space and undistorting them by warping. Then, the corners are identifyed by using Fofi et al.'s approach [22] in order to calculate the image homography between the camera image and the projector image space.

B. GPU-based Reconstruction

The reconstruction follows a GPU-based approach from Ladikos et al. [24], where silhouettes of the detected gasflow are downsampled so that every voxel projects into a single pixel. This way the projection of the voxel center point suffices for the lookup in the silhouette images. Note, that the silhouettes need not necessarily be coherent regions of genus 0. In our case we simply performed an opening on the spot regions with sufficiently large optical flow to arrive at a binary mask that we used as silhouette input for the hull generation. The downsampling is done by Gaussian smoothing followed by a downscaling. The kernel is then executed for every voxel by deriving the 3D position from its id and by projecting the voxel center point into the image. Early rejection based on image occupancy is performed to ease computational burden.

VI. Discussion

A. Optical Flow Comparison

We compared the subpixel accurate sparse pattern-based distortion detection against Horn-Schunck [2], Lucas-Kanade [1] and Drulea-Nedevschi [3] with a synthetic spot image and the same image modulated by the Groove2-owmap from the Middlebury database [25], that have been altered with varying noise and increasing amount of blurriness. We evaluated the average angular error (AAE) compared to the ground truth and found that the proposed method shows a lower AAE with increasing image degradation compared to the other flow algorithms, Fig. 4. Thus, the proposed algorithm is more robust against the noise introduced by the capturing setup.

B. Gas flow and occluder segmentation

The occluders have been segmented out by removing the overexposed regions in the input image. The projected area of the gas flow could be determined accurately to subpixel level in the input images for both the original IR camera image and the warped projector image.

C. GPU-based reconstruction

The reconstruction is written in C for CUDA and performs in 52.76ms at 64^3 voxel, in 417.79 ms 128^3 voxel and in 3033.89 ms at 256^3 voxel on an Intel Corei7-3960X (Six Core Extreme, 15MB Cache) Overclocked up to 4.0Ghz with a 4GB GDDR5 NVIDIA GeForce GTX 690 graphics card. Significant gain in performance can be achieved by using an octree-based approach. An exemplaric visualization of the reconstructed volumes is depicted in Fig. 5 for different occluders: We captured a sequence of 39 frames with three Microsoft Kinects and processed it as proposed. It can be seen that the droplet occluder (first row) smoothly introduces into the flow, while the golfball (second row) forces the gas to direct leftwards. The bridge occluder obstructs the gas flow and forces the flow to be inverted downwards.

VII. Conclusion

We presented a new method for subpixel accurate flow detection for sparse image data such as the Microsoft Kinect spot pattern. We found that the discretized skew Gaussian approach outperforms traditional and current optical flow approaches in the context of the Kinect's IR image stream. We tested the algorithm in a multiple Kinect vision architectur to detect the interface of propane flow around occluding objects in air. In that use case scenario we sought to reconstruct the geometric extend of gas flows under the presence of occluders captured with three Kinect placed in an half-arc around the flow with projection walls placed at a fixed distance opposite to each Kinect. We exploited the sparse spot detection algorithm to provide masks for a

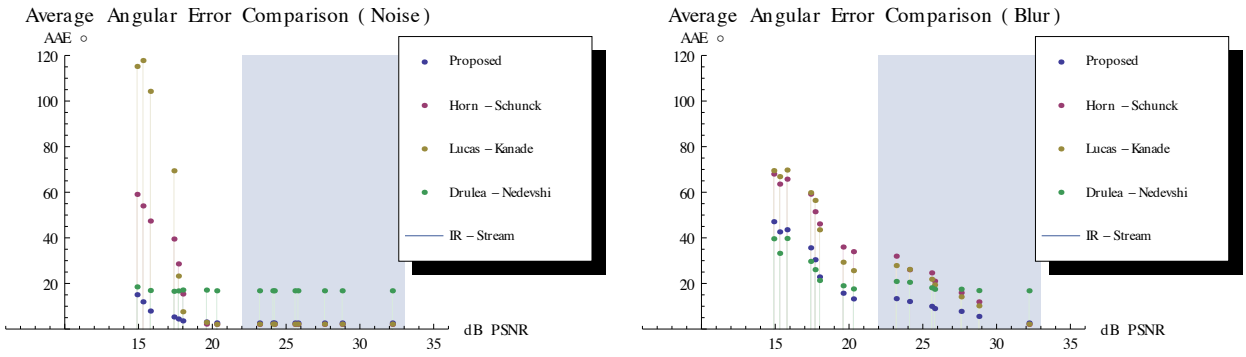

Fig. 4. Comparison of state of the optical flows to the proposed spot-based optical flow for varying levels of noise (left) and blur (right) applied to a synthetic spot pattern. The ground truth optical flow is generated with the *groove2* dataset from the Middlebury database [25]. The proposed approach (blue) outperforms the state-of-the art approaches in the range relevant for Kinect data stream processing (light blue region).

GPU-based visual hull reconstruction. By incorporating the projector extrinsics we doubled the reprojection information in order to increase accuracy. In the future we seek to provide for a real-time solution consisting of multiple Kinects in the context of the Kinect at home project [26], that has been recently launched. The goal would be to provide for realtime detection of gas leakages or heat dissipation within the living room environment.

ACKNOWLEDGMENT

Supplementary Videos are accessible under
https://dl.dropbox.com/u/21912442/supplementary.zip

REFERENCES

[1] B. D. Lucas and T. Kanade (1981), An iterative image registration technique with an application to stereo vision. Proceedings of Imaging Understanding Workshop, pages 121–130

[2] B.K.P. Horn and B.G. Schunck, "Determining optical flow." Artificial Intelligence, vol 17, pp 185-203, 1981

[3] Total variation regularization of local-global optical flow Marius Drulea, Sergiu Nedevschi (accepted to ITSC 2011)

[4] The Multivariate Skew-Normal Distribution A Azzalini, A Dalla Valle - Biometrika, 1996 - Biometrika Trust

[5] Atcheson B., Heidrich W., Ihrke I.: An evaluation of optical flow algorithms for background oriented schlieren imaging. Experiments in fluids 46, 3 (2009), 467476.

[6] Atcheson B., Ihrke I., Heidrich W., Tevs A., Bradley D., Mag- nor M., Seidel H.: Time-resolved 3d capture of non-stationary gas flows. In ACM Transactions on Graphics (TOG) (2008), vol. 27, ACM, p. 132.

[7] Barron J., Fleet D., Beauchemin S.: Performance of optical flow techniques. International journal of computer vision 12, 1 (1994), 43 77.

[8] Berger K., Ihrke I., Atcheson B., Heidrich W., Magnor M., et al.: Tomographic 4d reconstruction of gas flows in the presence of occluders. In Vision, Modeling, and Visualization Workshop (VMV) (2009), pp. 18.

[9] Bown M., MacInnes J., Allen R., Zimmerman W.: Three- dimensional, three-component velocity measurements using stereo- scopic micro-piv and ptv. Measurement Science and Technology 17 (2006), 2175.

[10] Berger K., Ruhl K., Albers M., Schroder Y., Scholz A., Guthe S., Magnor M.: The capturing of turbulent gas flows using multiple kinects. In Proc. CDC4CV 2011 (Nov. 2011), IEEE. ISBN: 978-1- 4673-0061-2.

[11] Baker S., Scharstein D., Lewis J., Roth S., Black M., Szeliski R.: A database and evaluation methodology for optical flow. Interna- tional Journal of Computer Vision 92, 1 (2011), 131.

[12] Bendicks C., Tarlet D., Roloff C., Bordas R., Wunderlich B., Michaelis B., Thevenin D.: Improved 3-d particle tracking ve- locimetry with colored particles. Journal of Signal and Information Processing 2, 2 (2011), 5971.

[13] Cowen E., Monismith S., Cowen E., Monismith S.: A hybrid digital particle tracking velocimetry technique. Experiments in Fluids 22, 3 (1997), 199211.

[14] Czarske J.: Laser doppler velocimetry using powerful solid-state light sources. Measurement Science and Technology 17 (2006), R71.

[15] Drulea M., Nedevschi S.: Total variation regularization of local- global optical flow. In Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on (oct. 2011), pp. 318 323.

[16] Hinz S.: Fast and subpixel precise blob detection and attribution. In Image Processing, 2005. ICIP 2005. IEEE International Conference on (2005), vol. 3, IEEE, pp. III457.

[17] Horn B., Schunck B.: Determining optical flow. Articial intelli- gence 17, 1-3 (1981), 185203.

[18] Lucas B., Kanade T.: An iterative image registration technique with an application to stereo vision. In Proceedings of the 7th international joint conference on Articial intelligence (1981).

[19] Svoboda T., Martinec D., Pajdla T.: A convenient multicamera self-calibration for virtual environments. Presence: Teleoperators and Virtual Environments 14, 4 (2005), 407422.

[20] Ten Cate A., Nieuwstad C., Derksen J., Van den Akker H.: Particle imaging velocimetry experiments and lattice-boltzmann sim- ulations on a single sphere settling under gravity. Physics of Fluids 14 (2002), 4012.

[21] Willert C., Gharib M.: Digital particle image velocimetry. Experiments in fluids 10, 4 (1991), 181193.

[22] G. Falcao, N. Hurtos, J. Massich, D. Fofi, "Projector-Camera Calibration Toolbox", http://code.google.com/p/procamcalib, 2009.

[23] A multivariate skew normal distribution AK Gupta, G Gonzlez-Faras - Journal of Multivariate, 2004 - Elsevier

[24] A. Ladikos, S. Benhimane, N. Navab Efficient Visual Hull Computation for Real-Time 3D Reconstruction using CUDA IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska (USA), June 2008. Workshop on Visual Computer Vision on GPUs (CVGPU)

[25] A Database and Evaluation Methodology for Optical Flow, published open access in International Journal of Computer Vision, 92(1):1-31, March 2011.

[26] http://www.kinectathome.com/

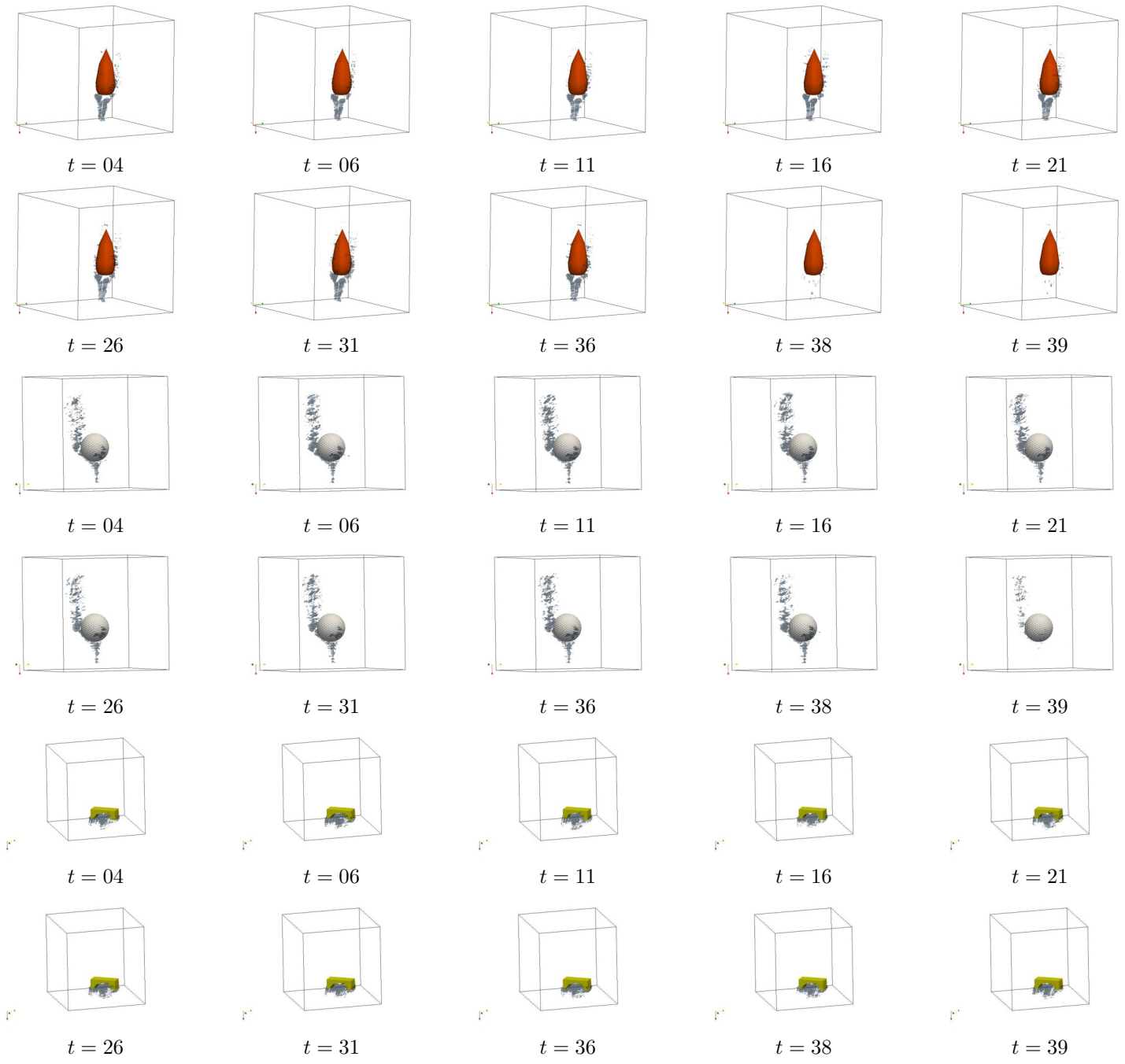

Fig. 5. A sequence of 39 frames captured with three Microsoft Kinects and processed with the proposed algorithm for propane gas flowing around obstructing objects with different aerodynamic properties. The voxel volumes have resolution 128^3, each input image roughly consists of 2000 spots in the region of interest. While the droplet occluder (first row) smoothly introduces into the flow, the golfball (second row) forces the gas to direct leftwards. The bridge occluder completely obstructs the gas flow, its flow direction is inverted downwards.