Using Active Illumination for Accurate Variational Space-Time Stereo

Sergey Kosov, Thorsten Thormählen, and Hans-Peter Seidel

Max-Planck-Institut Informatik (MPII), Saarbrücken, Germany

Abstract. This paper addresses the problem of space-time stereo with active illumination and presents a formulation of this problem in the variational framework. Variational problems of this scale are computationally expensive to solve directly. We overcome this challenge by showing that speed-improving techniques, as the full-multi-grid and the multi-level-adaptation techniques, can be applied. We evaluate the performance of our method on 3 ground-truth datasets. The experimental results for synthetic and real datasets show that the combination of active illumination and variational space-time stereo improves the quality of the reconstruction on average by up to 3.1 times compared to a reconstruction from a single passive stereo image pair without active illumination.

1 Introduction

A classical problem in computer vision is the reconstruction of disparity field between several stereo images. The task is to find those corresponding pixels in the stereo images that are the projections of the same 3D point. The collection of displacement values for all pixels of an image forms a dense disparity map.

Algorithms for dense disparity map reconstruction are often a basic building block of more complicated systems for automatic 3D scene analysis, event detection, or object recognition. These systems are applied, for example, in machine vision, robotics, or medical applications. Furthermore, the recent development in cinematography, where more movies are shot in stereo, has sparked a new interest in the topic of stereo estimation in academia as well as in the post-production and home-entertainment industry.

A number of researchers have worked on fast and accurate stereo estimation (e.g., [1,2,3]). Nowadays, variational methods are among the best techniques for optic flow reconstruction, which is very related to disparity estimation, e.g. Mémin and Pérez [4] and Brox et al. [5]. These methods minimize an energy functional by solving the corresponding Euler-Lagrange equation. In order to solve this equation numerically, it is represented as a system of parabolic partial differential equations in finite differences. To optimize the energy functional, iterative solvers, like the Jacobi and Gauss-Seidel methods, are used. In 2006, Bruhn et al. [6] presented a real-time implementation of a variational solver for optic flow reconstruction, based on the multigrid method [7]. Recently, Valgaerts et al. [8] have presented an approach that allows to incorporate both

A. Heyden and F. Kahl (Eds.): SCIA 2011, LNCS 6688, pp. 752–763, 2011.

spatial and temporal information (from two subsequent image pairs) for optic flow reconstruction. A real-time variational solver for disparity reconstruction was demonstrated [9]. The real-time performance has been achieved by a combination of the full-multi-grid (FMG) method, the multi-level adaptation technique (MLAT) [10], and adaptive parameter techniques.

Classical stereo vision algorithms process the stereo image pairs of different points in time independently. However, better results can be obtained by considering the problem not only in space but also in time. In 2003, Zhang et al. [11] showed that the space-time approach gives better results for disparity estimation. They suggested that each classical stereo algorithm can be extended to the spatio-temporal domain. In this paper we follow their suggestion and show how a variational solver can be used for space-time disparity estimation.

Nevertheless, though space-time approaches can improve disparity estimation results, almost all disparity estimation methods need local textures to compute dense disparity maps. Therefore, the algorithms lose their accuracy in homogeneous image regions. Additional texture can be generated when active illumination is applied [12]. In this paper it will be argued that projected vertical color strip pattern are very well suited to be used in combination with the variational method. Furthermore, infrared light can be used in order to project patterns that are not visible to the human eyes [13].

In this paper, we present a combination of structured light and fast variational space-time stereo. To the best of our knowledge, this paper is the first to perform space-time disparity estimation with active illumination in the variational framework. However, we are of course not the first to combine active illumination and space-time stereo. The benefit of additional spatio-temporal information for stereo vision has been shown before, e.g., by Zhang et al. [11]. The advantage of the variational framework is the high reconstruction accuracy. The disadvantage of variational solvers is that they tend to become slow, if they are applied on large equation systems. This is especially a problem as the number of equations is increased by adding information from different points in time. In this paper, we show that speed-improving techniques, like FMG and MLAT, can still be applied. The approach is evaluated with 3 ground-truth datasets.

2 The Space-Time Variational Method

Currently, almost all stereo vision algorithms analyze and process stereo image sequences in pairs of a left and a right frame. Processing the stereo pairs separately from the whole sequence leads to the loss of the dynamics occurring in this image sequence [14]. However, these dynamics (such as displacement between two frames, or occluded/exposed areas) contain vital information and can be used to achieve a better convergence rate for the variational method and to enhance the accuracy. We address this issue by extending the variational approach to the spatiotemporal domain, as described in the following subsections.

Problem formulation. Given a rectified stereo sequence consisting of individual sequences for the left and right camera, each scalar-valued image sequence

$I(x, y, t)$ is stored in a pixel matrix and $(x, y, t)^\top$ is the space-time coordinate of a voxel within the three-dimensional spatio-temporal domain $\bar{\Omega} = \Omega \times T$. For every voxel of the left sequence $I_l(x, y, t)$, we now try to estimate the disparity value $u(x, y, t)$, which is the offset of the x-coordinate of the voxel position, in order to match the corresponding voxel from the right sequence $I_r(x, y, t)$:

$$I_r(x, y, t) - I_l(x + u(x, y, t), y, t) = 0 \quad . \tag{1}$$

Since we are dealing with continuous real-world data, the disparities are not necessarily integer values. This is taken into account by employing different linearization techniques while solving Eq. (1). In our case, we use a linear interpolation approach [9] for the linearization.

An energy functional is constructed that consists of two terms: a data term that imposes the constancy assumption on the grey values, and a smoothness term that regularizes the local and often non-unique solution for the data term by an additional smoothness assumption.

Data term. In a real-world recording, we have to deal with occlusions and non-lambertian surfaces in the scene. Therefore, the left-hand side of Eq. (1) is usually not exactly zero. However, it should be as close to zero as possible. Therefore, we minimize the corresponding energy functional

$$E(u(x, y, t)) = \iiint_{\bar{\Omega}} \|I_r(x, y, t) - I_l(x + u(x, y, t), y, t)\|^2 \, dx \, dy \, dt \quad . \tag{2}$$

Smoothness term. The smoothness term is based on the assumption that neighboring space-time regions belong to the same object and, thus, have similar disparities. The main role of the smoothness term is the redistribution of the computed information and the elimination of local disparity outliers. If reliable information from the data term is not available, the smoothness term helps to fill the problematic region with disparities calculated from neighboring regions and from previous and future points in time.

In our work, we use 3 different regularizers: Tichonov, Charbonnier, and Perona-Malik regularization. Tichonov regularization assumes overall smoothness and does not adapt to semantically important image or disparity field structures (Horn and Schunck [15]). Charbonnier's and Perona-Malik's disparity-driven regularizations assume piecewise smoothness and respect discontinuities in the disparity field (see, e.g., [16,17]).

For all three regularizers, the smoothness term in general form is expressed as $\Psi(|\nabla_3 u(x, y, t)|^2)$. The energy functional from Eq. (2) extended by the smoothness term takes the following form:

$$E(u) = \iiint_{\bar{\Omega}} \|I_r(x, y, t) - I_l(x + u, y, t)\|^2 + \varphi \cdot \Psi(|\nabla_3 u|^2) \, dx \, dy \, dt \quad , \tag{3}$$

where φ is the weight of the smoothness term.

Euler-Lagrange equation. The goal of the variational method is to find a function $u(x, y, t)$, which minimizes the energy functional $E(u(x, y, t))$. Once we have constructed the energy functional, we need to find a solution, i.e., disparity field, which minimizes the functional. If the functional is constructed over a strict convexity requirement, the problem of minimization can be simplified, since there exists only one unique solution.

The Euler-Lagrange equation is an equation that is satisfied by the unknown function $u(x, y, t)$, which minimizes the functional

$$E(u)) = \iiint_{\bar{\Omega}} F(x, y, t, u, u_x, u_y, u_t)\, dx\, dy\, dt \quad , \tag{4}$$

where $u_x = \frac{\partial u}{\partial x}$, $u_y = \frac{\partial u}{\partial y}$, $u_t = \frac{\partial u}{\partial t}$ and F is a given function that has continuous first order partial derivatives. The Euler-Lagrange equation then is the partial differential equation:

$$F_u - \frac{\partial}{\partial x} F_{u_x} - \frac{\partial}{\partial y} F_{u_y} - \frac{\partial}{\partial t} F_{u_t} = 0 \quad . \tag{5}$$

For the energy functional from Eq. (3) the Euler-Lagrange equation for each voxel $(x, y, t)^\top$ is given by

$$I_{lx}(x + u, y, t)(I_r(x, y, t) - I_l(x + u, y, t)) + \varphi \cdot \mathrm{div}(\Psi'(|\nabla_3 u|^2) \cdot \nabla_3 u) = 0 \quad . \tag{6}$$

In order to minimize the energy functional, we solve the resulting system of differential equations with homogeneous Neumann boundary conditions [18]. This step is done via discrete numerical schemes. The Euler-Lagrange equations are discretized, linearized, and approximated via finite-differences schemes. In the end, we arrive at a linear (in case of Tichonov regularizer) or non-linear (in case of Charbonnier or Perona-Malik regularizers) system of equations.

Discretization. In order to discretize Eq. (6), we use linear interpolation for the data term and standard discretization for the diffusion filters [19]. For the space-time variational method we use a 6-voxel stencil (see Fig. 1, left) for the computation of the smoothness term (instead of a 4-pixel stencil as is used by classical variational approaches that do not smooth in time dimension). The discretized smoothness term can be written as $\Psi'(|\nabla_3 u(i\Delta x, j\Delta y, k\Delta t)|^2) \equiv$

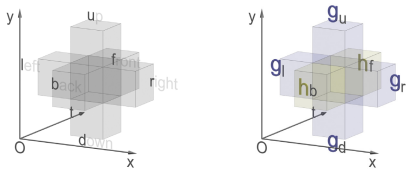

Fig. 1. 3D stencil for the smoothness term discretization: **left:** labeling of weighting coefficients; **right:** splitted stencil for different space and time regularization

$g_{i,j,k}$. We also introduce the following substitutions (cf. Fig. 1): $g_{i+1,j,k}+g_{i,j,k} \equiv g_r$, $g_{i-1,j,k}+g_{i,j,k} \equiv g_l$, $g_{i,j+1,k}+g_{i,j,k} \equiv g_u$, $g_{i,j-1,k}+g_{i,j,k} \equiv g_d$, $g_{i,j,k+1}+g_{i,j,k} \equiv g_f$, $g_{i,j,k-1}+g_{i,j,k} \equiv g_b$ and $\sum_{o\in\{l,r,u,d,f,b\}} g_o \equiv g_c$. Then the smoothness term takes the following expression:

$$\begin{aligned}\varphi \cdot [\ & g_r \cdot u(x+1,y,t) + g_l \cdot u(x-1,y,t) + g_u \cdot u(x,y+1,t) \\ & + g_d \cdot u(x,y-1,t) + g_f \cdot u(x,y,t+1) + g_b \cdot u(x,y,t-1) \\ & - g_c \cdot u(x,y,t)] \quad .\end{aligned} \tag{7}$$

Since *space* and *time* are incommensurable concepts in mathematics, we may want to penalize the solution in space and time differently (e.g., the Charbonnier regularizer for space, and the Tichonov regularizer for time). For that purpose we modify the standard discretization scheme and introduce the alternative function $h_{i,j,k}$, and parameter ϕ to function $g_{i,j,k}$ and parameter φ, respectively. Eq. 7 can be rewritten as

$$\begin{aligned}\varphi \cdot [\ & g_r \cdot u(x+1,y,t) + g_l \cdot u(x-1,y,t) + g_u \cdot u(x,y+1,t) \\ & + g_d \cdot u(x,y-1,t) - g'_c \cdot u(x,y,t)] \\ +\phi \cdot [\ & h_f \cdot u(x,y,t+1) + h_b \cdot u(x,y,t-1) - h'_c \cdot u(x,y,t)] \quad ,\end{aligned} \tag{8}$$

where $g'_c \equiv \sum_{o\in\{l,r,u,d\}} g_o$ and $h'_c \equiv \sum_{o\in\{f,b\}} h_o$.

Note that Eq. (8) will be identical to Eq. (7) for $h_{i,j,k} \equiv g_{i,j,k}$ and $\phi = \varphi$. Therefore Eq. (8) describes the more general case and provides the additional flexibility. In addition, for $\phi = 0$, we end up with the classical variational approach for disparity reconstruction.

Space-time FMG and MLAT. The multigrid method implies the usage of coarser grids, i.e., a pyramid of scaled versions of the initial images. Classical multigrids methods use the factor of two as a scale factor. So the maximal reasonable number of levels we can express are: $\#\text{levels} \leq log_2 \min\{X,Y,T\}$.

Usually, the time dimension is much smaller than the space dimensions: $T << \min\{X,Y\}$, e.g., if we have video at 25 fps and want to process several blocks per second, we could use for example around 10 frames in a space-time block. On the other hand, the image spatial resolution measures in hundreds of pixels. Therefore, the number of coarse grids will be too small for effective application. As a consequence, we have reverted to the solution to use the full-multi-grid approach only in the spatial directions for each time slice independently. Thus, for each frame, we use the same number of pyramid levels. Implementation details about the of the FMG approach in the variational framework are given in [9].

The multi-level adaptation technique (MLAT) is another technique to reduce the computation time of the variational solver. When updating from a coarser to finer grid in the FMG approach, we look at peculiarities of the solution, and only refine the grid in areas where high peculiarities are found. This results in a non-regular grid structure on the finest level. During space-time stereo processing it can be assumed that the solution is changing very smoothly with time: $u_t \to 0$. Thus, in order to apply the MLAT in the space-time framework, we can use the

classical MLAT approach for calculating the adapted grid once per currently processed space-time block. As the structure is the same for all frames of the block, the neighbouring voxels in time have the same spatial resolution, and thus can be directly used to perform the iterative variational optimization steps with the described 6-voxel stencil.

3 Active Illumination

In most stereo matching algorithms, the inherent ambiguity of image values in homogeneous image regions leads to a loss of accuracy in the computation of dense disparity maps. A possible solution to this problem is the introduction of artificial texture into the scene, e.g., by the projection of intensity coded light [12]. These stereo setups with active illumination benefit from improved local scene texture – hence better correspondences – and therefore allow the reconstruction of more accurate dense disparity maps.

Our setup consists of a Point Grey Bumblebee® XB3 camera synchronized with a projector casting a structured light pattern onto the scene (see Fig. 2). We projected patterns in the visible light spectrum but projecting infra-red patterns, which are not visible to the human eye, is possible as well.

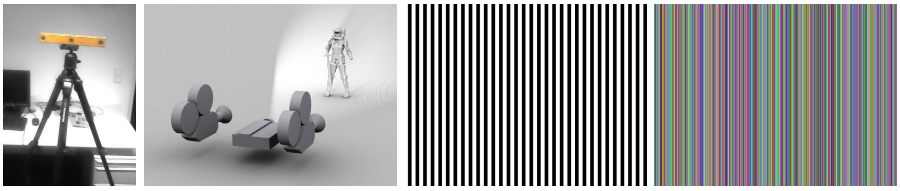

Fig. 2. Our stereo setup and examples of projected patterns: (from left to right: Point Grey Bumblebee® XB3 camera; principle of a stereo system using active illumination; binary pattern; random color pattern

The question is now what is the best suited projection pattern for the scene and for our reconstruction method. A number of researchers have worked on optimizing the employed patterns that are mostly based on vertical stripes [20,21]. The main idea of these works is to adapt the pattern in such a way, that interference by the scene is minimized. Using a graph-cuts approach, the algorithm identifies the pattern and reconstructs the disparity map. Thereby, a combination of geometric coding, color coding and tracking over time is used.

Using with the variational method and a stereo camera, we do not need to identify the pattern during the reconstruction process. That gives us the following advantages: 1.) we do not need to know the exact position of the projector and 2.) we do not need prior knowledge about the structure of the projected light. Let $I(x, y, t)$ be an image of a scene without active illumination used (only with ambient illumination), and $I^C(x, y, t)$ will be the image of the scene with active illumination used (plus the same ambient illumination) and taken from

the same position as $I(x, y, t)$. Then the $C(x, y, t) = I^C(x, y, t) - I(x, y, t)$ will be the visible color pattern.

The data term of the Euler-Lagrange equation (6), which we solve to estimate the disparity map, has the following form: $I_{lx}(I_r - I_l)$. In order to achieve accurate results, we need to make this term as informative as possible (for details, please refer to [9]). Using structured light, we can express this demand as follows: $\frac{\partial I^C}{\partial x} >> 0$, or

$$\arg\max_C \frac{\partial I^C}{\partial x} = \arg\max_C \frac{\partial (I + C)}{\partial x} = I_x + \arg\max_C C_x \quad , \tag{9}$$

i.e. the optimal structured pattern for the variational approach is such a pattern, where the horizontal derivative is maximal. Thus, vertical black-and-white stripes (see Fig. 2) without any adaptation to $I_x(x, y, t)$ or to the observed scene are best suited to achieve highest accuracy. This conclusion agrees with the work of Horn and Kiryati [22]. However, such a pattern has some important disadvantages. If the scene lacks texture, after illuminating it with the black and white stripes pattern, we get the same uncertainty, because of the repeating structure of the pattern. Moreover, using multi-grids, the variational method needs to have a set of coarser copies of the input images, and using the black-and-white stripes with geometric coding could cause ambiguities on some coarse levels. As a solution, we propose a random generated, high-contrast color pattern (see Fig. 2). Since this pattern is generated randomly, all the levels (coarse and fine) will contain random vertical lines based illumination.

Assuming all surfaces in the scene obey Lambertian reflection, the reflected light resulting from the blending of projected light and object texture is identical in both images [23].

4 Evaluation

In this section we show results for the presented space-time variational method with active illumination for disparity estimation. The results are also presented in the supplemental video. The method was implemented, using single-threaded C++ code, and all the experiments were made on a Intel Core 2 Quad Processor 2,83GHz with 8GB DDR2 RAM. To evaluate our novel method, we use three different data sets with ground-truth disparity and occlusion maps. We evaluate the results with the percentage of "bad" voxels (which have a disparity error larger than one or two pixels, see [24] for more details).

Dynamic scene without active illumination. As a first example, we use the data set *Gargoyle* from York University [25], with a resolution of 640x480x40 voxels. The results for one of the frames from the sequence for the classical variational single frame processing (SFP) and novel variational space-time processing (STP) approaches can be found in Fig. 3. For both approaches, the number of iterations at the finest level is 10, and the parameter λ of the non-linear

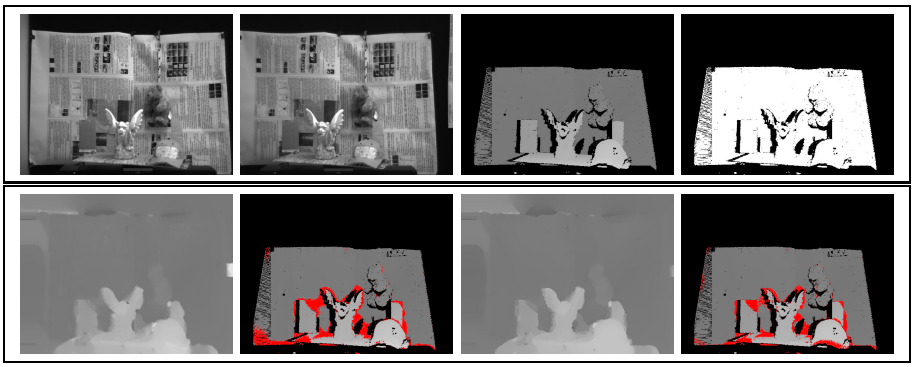

Fig. 3. Gargoyle scene: **top row** (from left to right): left and right images of the 27th stereo pair from the sequence, the corresponding ground truth disparity and occlusion map; **bottom row:** (from left to right) solution and bad pixels map (9.69%) for the SFP approach, solution and bad pixels map (7.99%) for the STP approach (error threshold = 2 pixels) (red pixels = bad pixels).

Charbonnier regularizer is $\lambda = 0.03$. The only difference are the smoothness parameters, which are $\varphi = 2500$ for SFP, and $\varphi = 2300$ and $\psi = 100$ for STP.

The calculation time for the whole space-time block containing all 40 frames is 114 seconds. For the 27th stereo frame, we can observe an improvement from 9.69% for SFP to 7.99% for STP. Fig. 5 shows the percentage of bad pixels for the whole sequence. The STP curve is smoother than the SFP curve and on average we gain about 1.3% improvement (SFP: 10.42%; STP: 9.16%).

Static scene with active illumination. In a second experiment, we use the dataset *Ship*, with a resolution of 600x400x10 voxels. The scene was shot with a point Grey Bumblebee® XB3 camera and the ground truth was obtained with a Konika Minolta 3D laser scanner. In Fig. 4, the results for the classical SFP and novel STP are shown. For both approaches we set the number of iterations to 25, and used the linear Tichonov penalizer for time dimension and the non-linear Charbonnier and Perrona-Malik penalizers with $\lambda = 0.01$ and $\lambda = 0.3$, respectively, for space dimensions. The SFP approach is applied with smoothness parameter $\varphi = 5000$ and STP with $\varphi = 1000$ and $\psi = 1000$. The calculation time for this scene is 55 seconds.

For the first stereo frame, we can observe a large improvement from 14.74% for SFP to 10.41% for STP. This is because in static scenes the solution is the same for all the frames of the input sequence, and therefore the algorithm can rely strongly on the temporal regularizer and we can use a very sharp penalizer for the space dimensions. As a result, we gain more accurate edges and less noise in the background of the reconstructed disparity map. Fig. 5(right) shows the percentage of bad pixels for the whole sequence. We can observe that the STP curve is almost constant and concurs with the average value of bad voxels. In contrast, SFP curve is not smooth and on average we gain an improvement of about 4% (SFP: 14.56%; STP: 10.54%).

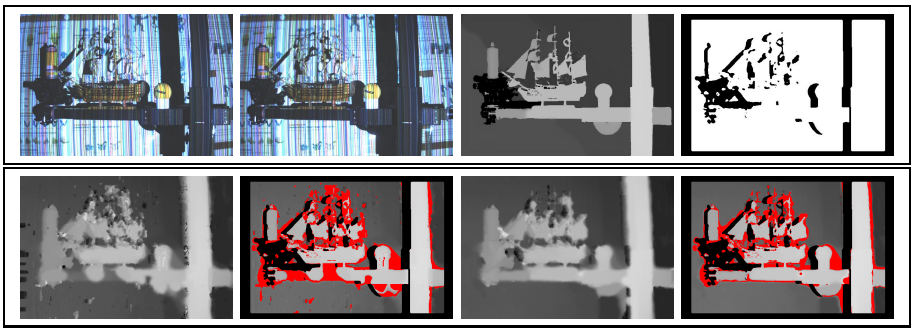

Fig. 4. Ship scene: **top row** (from left to right): left and right images of the first stereo pair from the sequence, the ground-truth disparity and occlusion map; **bottom row:** (from left to right) solution and bad pixels map (14.74%) for the SFP approach, and solution and bad pixels map (10.41%) for the STP approach (error threshold = 2 pixels) (red pixels = bad pixels).

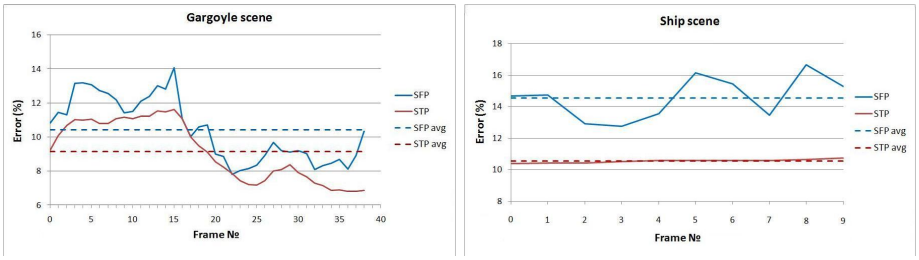

Fig. 5. Comparison of the reconstruction accuracy for SFP and STP (error threshold = 2 pixels). **left:** Gargoyle scene: Average values: SFP: (10.42%); STP: (9.16%); **right:** Ship scene: Average values: SFP: (14.56%); STP: (10.54%).

Dynamic scene with active illumination. The last dataset is the synthetic scene *Knight*, with a resolution of 640x360x18 voxels. The 3D scene and ground truth data are generated in manually in a 3D modeling package. The scene has a static background and a rotating knight shell. The maximal disparity in the scene is 16 pixels. In Fig. 6 the results for the novel STP approach with and without active illumination is shown. For both cases we used 15 iterations on the finest level and non-linear Charbonnier and Perrona-Malik penalizers with $\lambda = 0.01$ and $\lambda = 0.3$ for time and space dimensions, respectively. The smoothness parameters are $\varphi = 2500$ and $\psi = 1100$.

For the 7th stereo frame (shown in Fig. 6), we can observe more than two times improvement, from 4.46% of bad pixels without active illumination to 2.13% with active illumination. As we can see from the Fig. 6, with active illumination and space-time processing it became possible to completely get rid of bad pixels in the static background, and significantly reduce the amount of bad pixels at the edges of the moving objects due to the application of the non-linear regularization in the time dimension. The processing of the whole scene took 58 seconds. The

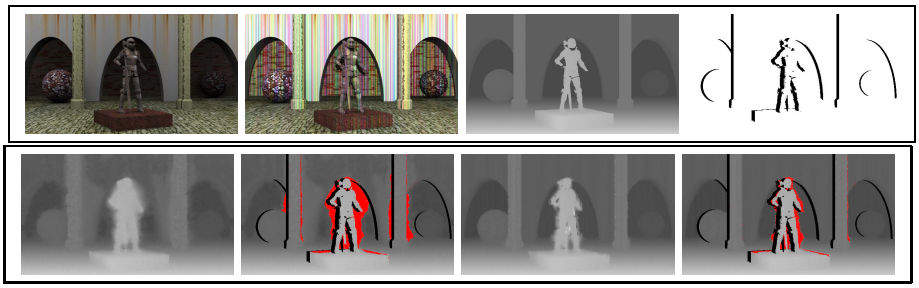

Fig. 6. Knight scene: **top row** (from left to right): left images of the 7th stereo pair from the sequence (without active illumination (AI) and with AI), the ground truth, and occlusion map; **bottom row:** (from left to right) solution and bad pixels map (4.46%) for STP approach without AI, and solution and bad pixels map (2.13%) for STP with AI (error threshold = 1 pixel) (red pixels = bad pixels)

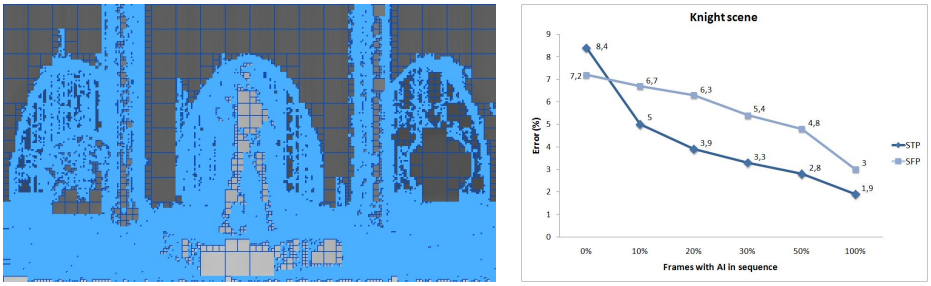

Fig. 7. **left:** The finest MLAT grid for the *Knight* scene; **right:** Emitting active illumination color pattern each i-th frame (error threshold = 1 pixel)

finest grid, calculated by MLAT and used for all 18 frames of the sequence is depicted in Fig. 7(left).

We further evaluated the approach by applying active illumination not all the time, but only each i-th frame. This can significantly reduce the energy consumption for an LED-structured light projector in real-life applications without losing too much accuracy in reconstructed disparity maps. The results of this experiment are shown in Fig. 7(right). We used the following parameters: number of iterations is 20, non-linear Charbonnier and Perrona-Malik penalizers with $\lambda = 0.01$ and $\lambda = 0.3$ for time and space dimensions, respectively, and smoothness parameters $\varphi = 2300$ and $\psi = 1500$.

From the diagrams in Fig. 7(right) we can observe that if there are no frames with active illumination used in the sequence, the SFP approach gives better results than STP. That is because we used the same parameters for the whole experiment, and these parameter must be a trade-off between processing scenes with active illumination and without. In our case, STP gives worse results, because of too strong smoothness of the solution in time direction, which resulted in a too strong blurring of the disparity map around the moving objects in scene. On the other hand, we can observe that already with 20% frames with active

Table 1. Comparison of the average percentage of bad pixels for the *Knight* scene (error threshold = 1 pixel). Improvement ratio are given in brackets.

	without AI		with AI	
	SFP	STP	SFP	STP
Variational method	5.9%	5.2% (1.1x)	3% (**2**x)	1.9% (**3.1**x)
Expansion method	2.9%	3.7% (0.8x)	2.4% (1.2x)	2.2% (1.3x)
Belief propagation	3.1%	2.9% (1.1x)	1.9% (1.6x)	1.4% (2.2x)
Swap method	10.3%	8.8% (1.2x)	7.2% (1.4x)	4% (2.6x)
Infection method	16.3%	16.2% (1.0x)	9.1% (1.8x)	8.9% (1.8x)
TRW method	3.6%	2.4% (**1.5**x)	2.4% (1.5x)	1.3% (2.8x)

illumination used in the sequence we observe two times better accuracy than with SFP.

In the Tab. 1 we show the best results that are gained with different combinations of the proposed variational methods in the top row of the table. The worst case is the processing of each frame separately and without active illumination. The best result was achieved for the space-time approach with active illumination with 1.9% of bad voxels. The other rows of Tab. 1 show results for classical methods, which implementations are available online[1]. The STP results of these methods are generated by smoothing the SFP results over time with the Tichonov regularizer.

5 Conclusion

We have shown that processing time-space blocks instead of single stereo image pairs provides significantly higher accuracy for the disparity reconstruction. Processing the time-space blocks within the variational method and the combination with active illumination has shown to be an effective approach. The accuracy of static scenes with different randomly generated color patterns increases the accuracy of reconstruction. A comparison on datasets with a classical variational disparity estimator, which processes only a single frame, shows that our implementation of the extended variational approach outperforms the current state-of-the-art. Furthermore, we showed how speed-improving techniques, like the full-multi-grid technique and the multi-level-adaptation technique, can be applied in the space-time stereo variational framework.

References

1. Ross, W.P.: A practical stereo vision system. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 148–153 (1993)
2. Kolmogorov, V., Zabih, R.: Computing visual correspondence with occlusions via graph cuts. In: International Conference on Computer Vision, pp. 508–515 (2001)
3. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: Proc. Computer Vision and Pattern Recognition, vol. I, pp. 195–202 (2003), `http://vision.middlebury.edu/stereo/`
4. Mémin, E., Pérez, P.: Dense estimation & object-based segmentation of the optical flow with robust techniques. IEEE Trans. on Image Processing 7, 703–719 (1998)

[1] Middlebury stereo evaluation web-site: http://vision.middlebury.edu/stereo/

5. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
6. Bruhn, A., Weickert, J., Kohlberger, T., Schnörr, C.: A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. Int. J. Comput. Vision 70, 257–277 (2006)
7. Fedorenko, R.: Relaxation method for solving elliptic differential equations. Journal of Computational Mathematics and Mathematical Phisics 1, 922–927 (1961)
8. Valgaerts, L., Bruhn, A., Zimmer, H., Weickert, J., Stoll, C., Theobalt, C.: Joint estimation of motion, structure and geometry from stereo sequences. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 568–581. Springer, Heidelberg (2010)
9. Kosov, S., Thormählen, T., Seidel, H.-P.: Accurate real-time disparity estimation with variational methods. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Kuno, Y., Wang, J., Wang, J.-X., Wang, J., Pajarola, R., Lindstrom, P., Hinkenjann, A., Encarnação, M.L., Silva, C.T., Coming, D. (eds.) ISVC 2009. LNCS, vol. 5875, pp. 796–807. Springer, Heidelberg (2009)
10. Brandt, A.: Multi-level adaptive technique (MLAT) for fast numerical solution to boundary value problems. Lecture Notes in Physics 18, 82–89 (1973)
11. Zhang, L., Curless, B., Seitz, S.: Spacetime stereo: Shape recovery for dynamic scenes. In: IEEE Conf. on Comp. Vision and Pattern Recognition, pp. 367–374 (2003)
12. Kang, S.B., Webb, J., Zitnick, C., Kanade, T.: A multibaseline stereo system with active illumination and real-time image acquisition. In: Proceedings of the Fifth International Conference on Computer Vision (ICCV 1995), pp. 88–93 (1995)
13. Frueh, C., Zakhor, A.: Capturing $2\frac{1}{2}$d depth and texture of time-varying scenes using structured infrared light. In: Proc. 3DIM, pp. 318–325 (2005)
14. Ristivojevic, M., Konrad, J.: Space-time image sequence analysis: Object tunnels and occlusion volumes. IEEE Transactions on Image Processing 15, 364–376 (2006)
15. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artificial Intelligence 17, 185–203 (1981)
16. Charbonnier, P., Aubert, G., Blanc-Ferraud, M., Barlaud, M.: Two deterministic half-quadratic regularization algorithms for computed imaging. In: International Conference on Image Processing, vol. 2, pp. 168–172 (1994)
17. Cohen, I.: Nonlinear variational method for optical flow computation. In: Eighth Scandinavian Conference on Image Analysis, vol. 1, pp. 523–530 (1993)
18. Cheng, A., Cheng, D.T.: Heritage and early history of the boundary element method. Engineering Analysis with Boundary Elements 29, 268–302 (2005)
19. Kosov, S.: 3D map reconstruction with variational methods. Master thesis, Saarland University (2008)
20. Blake, A., McCowen, D., Lo, H.R., Lindsey, P.J.: Trinocular active range-sensing. IEEE Trans. Pattern Anal. Mach. Intell. 15, 477–483 (1993)
21. Koninckx, T., Gool, L.V.: Real-time range acquisition by adaptive structured light. IEEE Transac. on Pattern Analysis & Machine Intelligence 28, 432–445 (2006)
22. Horn, E., Kiryati, N.: Toward optimal structured light patterns. In: Proc. 3DIM, p. 28. IEEE Computer Society, Washington, DC, USA (1997)
23. Koschan, A., Rodehorst, V., Spiller, K.: Color stereo vision using hierarchical block matching and active color illumination. In: ICPR 1996, vol. I, pp. 835–839 (1996)
24. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense stereo correspondence algorithms. International Journal of Computer Vision 47, 7–42 (2001)
25. Sizintsev, M., Wildes, R.P.: Spatiotemporal stereo via spatiotemporal quadric element (stequel) matching. In: CVPR, pp. 493–500 (2009)