Video Super Resolution using Duality Based TV-L^1 Optical Flow

Dennis Mitzel[1,2], Thomas Pock[3], Thomas Schoenemann[1] Daniel Cremers[1]

[1] Department of Computer Science
University of Bonn, Germany

[2] UMIC Research Centre
RWTH Aachen, Germany

[3] Institute for Computer Graphics and Vision
TU Graz, Austria

Abstract. In this paper, we propose a variational framework for computing a superresolved image of a scene from an arbitrary input video. To this end, we employ a recently proposed quadratic relaxation scheme for high accuracy optic flow estimation. Subsequently we estimate a high resolution image using a variational approach that models the image formation process and imposes a total variation regularity of the estimated intensity map. Minimization of this variational approach by gradient descent gives rise to a deblurring process with a nonlinear diffusion. In contrast to many alternative approaches, the proposed algorithm does not make assumptions regarding the motion of objects. We demonstrate good experimental performance on a variety of real-world examples. In particular we show that the computed super resolution images are indeed sharper than the individual input images.

1 Introduction

Increasing the resolution of images. In many applications of Computer Vision it is important to determine a scene model of high spatial resolution as this may help – for example – to identify a car licence plate in surveillance images or to more accurately localize a tumor in medical images. Fig. 1 shows a super resolution result computed on a real-world surveillance video using the algorithm proposed in this paper. Clearly, the licence plate is better visible in the computed superresolved image than in the original input image.

The resolution of an acquired image depends on the acquisition device. Increasing the resolution of the acquisition device sensor is one way to increase the resolution of acquired images. Unfortunately, this option is not always desirable as it leads to substantially increased cost of the device sensor. Moreover, the noise increases when reducing the pixel size.

Alternatively, one can exploit the fact that even with a lower-resolution video camera running at 30 frames per second, one observes projections of the same image structure around 30 times a second. The algorithmic estimation of a high

Fig. 1. In contrast to the upscaled input image [21] (left) the super resolution image computed with the proposed algorithm (right) allows to better identify the licence plate of the car observed in this surveillance video.

resolution image from a set of low resolution input images is referred to as *Super Resolution*.

General model of super resolution. The challenge in super resolution is to invert the image formation process which is typically modeled by series of linear transformations that are performed on the high resolution image, presented in Fig. 2.

Given N low resolution images $\{I_L^k\}_{k=1}^N$ of size $L_1 \times L_2$. Find a high resolution image I_H of size $H_1 \times H_2$ with $H_1 > L_1$ and $H_2 > L_2$ which minimizes the cost function:

$$E(I_H) = \sum_{k=1}^{N} \left\| P_k(I_H) - I_L^{(k)} \right\| \tag{1}$$

where $P_k(I_H)$ is the projection of I_H onto coordinate system and sampling grid of image I_L^k. $\|\cdot\|$ - can be any norm, but usually it is L^1 or L^2-norm. P_k is usually modeled by four linear transformations, that subject the high resolution image I_H to motion, camera blur, down sampling operations and finally add additive noise to the resulted low resolution image. Fig. 2 illustrates this projection. This projection connecting the k^{th} low resolution image to the high resolution image can be formulated using matrix-vector notation. [10]:

$$I_L^k = D_k B_k W_k I_H + e_k \tag{2}$$

where D_k is a down sampling matrix, B_k blurring matrix, W_k warping matrix and e_k a noise vector.

We use matrix-vector notation only for the analysis, the implementation will be realized by standard operations such as convolution, warping, sampling [10].

Fig. 2. The inversion of this image formation process is referred to as *Superresolution*.

Related Work. Super resolution is a well known problem and extensively treated in the literature. Tsai and Huang [9] were first who addressed the problem of recovering a super resolution image from a set of low resolution images. They proposed a frequency domain approach, that works for band limited and noise-free images. Kim et al. [8] extended this work to noisy and blurred images. Approaches in frequency domain are computationally cheap, but they are sensitive to model deviations and can only be used for sequences with pure global translational motion [7]. Ur and Gross proposed a method based on multi channel sampling theorem in spatial domain [6]. They perform a non-uniform interpolation of an ensemble of spatially shifted low resolution pictures, followed by deblurring. The method is restricted to global 2D translation in the input images. A different approach was suggested by Irani and Peleg [5]. Their approach is based on the iterative back projection method frequently used in computer aided tomography. This method has no limits regarding motion and handles non-uniform blur function, but assumes motion and blurring to be known precisely. Elad and Feuer proposed an unified methodology that combines the three main estimation tools in the single image restoration theory (ML) estimator, (MAP) estimator and the set theoretic approach using POCS [4]. The proposed method is general but assumes explicit knowledge of the blur and the smooth motion constraints.

In our approach we don't constrain to any motion model and don't assume the motion as to be known.

In recently published super resolution algorithms [17] and [16] the authors describe a super-resolution approach with no explicit motion estimation that is based on the Nonlocal-Means denoising algorithm. The method is practically limited since it requires very high computational power.

Contribution of this work In this paper we will present an robust variational approach for super resolution using L^1 error norm for data and regularization term. Rather than restricting ourselves to a specific motion model, we will employ a recently proposed high accuracy optic flow method which is based on quadratic relaxation [13]. We assume blur as space invariant and constant for all measured images, which is justified since the same camera was used for taking the video sequence.

This paper is organized as follows. In Section 2 we briefly review the optic flow estimation scheme introduced in [13]. In Section 3 we present super resolution approaches using L^2 and L^1 error norms for data and regularization terms. Subsequently, we present experimental results achieved with respective approaches. We conclude with a summary and outlook.

2 Optical Flow

The major difficulty in applying the above super resolution approach Eq. (2) is that the warping function W_k is generally not known. Rather than trying to simultaneously estimate warping and a super resolved image (which is computationally difficult and prone to local minima), we first separately estimate the

warping using the optic flow algorithm recently introduced in [13] and in the second step we use the motion to compute the inverse problem (2).

In this section we will shortly describe the optical flow algorithm as posed in [13]. An extension of this approach [2] was recently shown to provide excellent flow field estimates on the well-known Middlebury benchmark.

Formal Definition. Given two consecutive frames I_1 and $I_2 : (\Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R})$ of an image sequence. Find displacement vector field $\boldsymbol{u} : \Omega \rightarrow \mathbb{R}^2$ that maps all points of the first frame onto their new location in the second frame and minimizes following error criterion:

$$E(\boldsymbol{u}(\boldsymbol{x})) = \int_{\Omega} \lambda \left| I_1(\boldsymbol{x}) - I_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x})) \right| + (|\nabla \boldsymbol{u}_1(\boldsymbol{x})| + |\nabla \boldsymbol{u}_2(\boldsymbol{x})|) d\boldsymbol{x} \qquad (3)$$

where the first term (data term) is known as the optical flow constraint. It assumes that the grey values of pixels do not change by the motion, $I_1(\boldsymbol{x}) = I_2(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}))$. The second term (regularization term) penalizes high variations in $\boldsymbol{u}$ to obtain smooth displacement fields. λ weights between the both assumption.

At first we use the first order Taylor approximation for I_2 i.e.: $I_2(\boldsymbol{x} + \boldsymbol{u}) = I_2(\boldsymbol{x} + \boldsymbol{u}_0) + \langle (\boldsymbol{u} - \boldsymbol{u}_0), \nabla I_2 \rangle$ where $\boldsymbol{u}_0$ is a fix given disparity field. Since we linearized I_2, we will use multi-level *Coarse-to-Fine* warping techniques in order to allow large displacements between the images and to avoid trapping in local minima. Inserting the linearized I_2 in the functional (3) results in:

$$E(\boldsymbol{u}) = \int_{\Omega} \left[\lambda \left| I_2(\boldsymbol{x} + \boldsymbol{u}_0) + \langle (\boldsymbol{u} - \boldsymbol{u}_0), \nabla I_2 \rangle - I_1(\boldsymbol{x}) \right| + \sum_{d=1}^{2} |\nabla u_d(\boldsymbol{x})| \right] d\boldsymbol{x} \quad (4)$$

In the next step we label $I_2(\boldsymbol{x} + \boldsymbol{u}_0) + \langle (\boldsymbol{u} - \boldsymbol{u}_0), \nabla I_2 \rangle - I_1(\boldsymbol{x})$ as $\rho(\boldsymbol{u})$. We introduce additionally an auxiliary variable $\boldsymbol{v}$, that is a close approximation of $\boldsymbol{u}$ in order to convexify the functional and propose to minimize the following convex approximation of the functional (4):

$$E(\boldsymbol{u}, \boldsymbol{v}) = \int_{\Omega} \left[\sum_{d=1}^{2} |\nabla u_d| + \frac{1}{2\theta} (u_d - v_d)^2 + \lambda |\rho(\boldsymbol{v})| \right] d\boldsymbol{x} \qquad (5)$$

where θ is a small constant, such that v_d is a close approximation of u_d. This convex functional can be minimized alternately by holding $\boldsymbol{u}$ or $\boldsymbol{v}$ fix.

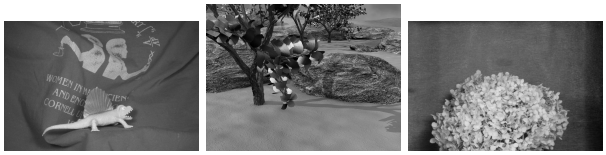
For a fix v_1 and $d = 1$ solve

$$\min_{u_1} \int_{\Omega} \frac{1}{2\theta} (u_1 - v_1)^2 + |\nabla u_1| \, d\boldsymbol{x} \qquad (6)$$

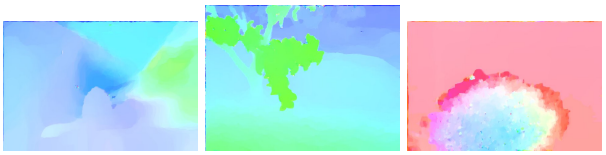
This is the denoising model that was presented by Rudin, Osher and Fatemi in [11]. An efficient solution for this functional was proposed in [1], which uses a dual formulation of (6) to derive an efficient and globally convergent scheme as shown in Theorem 1.

Theorem 1. *[1, 13] The solution for Eq. (6) is given by* $u_1 = v_1 + \theta \, \mathbf{div} \boldsymbol{p}$ *where* $\boldsymbol{p}$ *fulfils* $\nabla \left(v_1 + \theta \, \mathbf{div} \boldsymbol{p} \right) = \boldsymbol{p} \left| \nabla (v_1 + \theta \, \mathbf{div} \boldsymbol{p}) \right|$ *that can be solved by using semi-implicit gradient descent algorithm that was proposed by Chambolle [1]:*

$$\boldsymbol{p}^{n+1} = \frac{\boldsymbol{p}^n + \frac{\tau}{\theta} (\nabla (v + \theta \, \boldsymbol{div} \, \boldsymbol{p}^n))}{1 + \frac{\tau}{\theta} |\nabla (v + \theta \, \boldsymbol{div} \, \boldsymbol{p}^n)|} \qquad (7)$$



(a) Dimetrodon (b) Grove2 (c) Hydrangea



(d) AAE 6.8° (e) AAE 6, 2° (f) AAE 5.5°

Fig. 3. Performance evaluation of the test data from [3]. The first row shows a image from the input sequence. The second row shows the results obtained by implementation of the above algorithm by setting the parameters as follows $\lambda = 80.0$, $\theta = 0.4$ and $\tau = 0.249$.

where $\frac{\tau}{\theta}$ *is the time step,* $\boldsymbol{p}^0 = \boldsymbol{0}$ *and* $\tau \leq 1/4$.

The minimization for fix v_2 and $d = 2$ can be done in analogical.
For $\boldsymbol{u}$ being fixed, our functional (5) reduces to

$$E(\boldsymbol{v}) = \int_\Omega \left[\frac{1}{2\theta} \sum_{d=1}^{2} (u_d - v_d) + \lambda |\rho(\boldsymbol{v})| \right] d\Omega \tag{8}$$

Theorem 2. *[1, 13] The solution for the optimization problem (8) is given by the following threshold scheme:*

$$\boldsymbol{v} = \boldsymbol{u} + \begin{cases} \lambda\theta\nabla I_2 & \text{if} \quad \rho(\boldsymbol{u}) < -\lambda\theta|\nabla I_1|^2 \\ -\lambda\theta\nabla I_2 & \text{if} \quad \rho(\boldsymbol{u}) > \lambda\theta|\nabla I_2|^2 \\ -\rho(\boldsymbol{u})\nabla I_2/|\nabla I_2|^2 & \text{if} \quad |\rho(\boldsymbol{u})| \leq \lambda\theta|\nabla I_2|^2 \end{cases} \tag{9}$$

Theorem 2 can easily be shown by analyzing the cases $\rho(\boldsymbol{v}) < 0$, $\rho(\boldsymbol{v}) > 0$ and $\rho(\boldsymbol{v}) = 0$ – see [1, 13] for details.

Implementation. We implement the complete algorithm on the GPU by using CUDA framework and reached a high performance compared to the implementation on the CPU. Precise initialization of parameters $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{p}_d$ and further details can be found in [13]. Results that we could obtain with this approach are presented in Fig. 3.

3 Super Resolution

In this section we present variational formulations for motion-based super-resolution using first an L^2 norm and subsequently a more robust L^1 approach.

Variational Superresolution

In the first step we extend the data term which imposes similarity of the desired high-resolution image I_H (after warping W, blurring B and downsampling D) with the N observed images $\{I_L^k\}_{k=1}^N$ shown in Eq. (2) by a regularization term which imposes spatial smoothness of the estimated image I_H. To this end we start by penalizing the L_2 norm of its gradient [20]:

$$E(I_H) = \frac{1}{2}\sum_{k=1}^{N}\int_{\Omega}\left|D_k B_k W_k I_H - I_L^{(k)}\right|^2 + \lambda\left|\nabla I_H\right|^2 d\boldsymbol{x} \tag{10}$$

The regularization term is necessary since without it the inverse problem is typically ill-posed, i.e., does not possess a unique solution that depends continuously on the measurements. The parameter λ allows to weight the relative importance of the regularizer.

For finding the solution I_H, we minimize the energy function (10) by solving the corresponding Euler-Lagrange equation:

$$\frac{dE}{dI_H} = \sum_{k=1}^{N} W_k^\top B_k^\top D_k^\top (D_k B_k W_k I_H - I_L^{(k)}) - \lambda \triangle I_H = 0 \tag{11}$$

The linear operators $D_K^\top$, $W_k^\top$ and $B_k^\top$ denote the inverse operations associated with the down-sampling, warping and blurring in the image formation process. Specifically, $D_k^\top$ is implemented as a simple up-sampling without interpolation. $B_k^\top$ can be implemented by using the conjugate of the kernel: If $h(i,j)$ is the blur kernel then the conjugate kernel $\tilde{h}$ satisfies for all i, j: $\tilde{h}(i,j) = h(-i,-j)$. In our approach we model blurring through a convolution with an isotropic Gaussian kernel. Since the Gaussian kernel h is symmetric , the adjoint kernel $\tilde{h}$ coincides with h. In addition, we will assume that blurring and downsampling is identical for all observed images such that we can drop the index k in the operators B and D. The operator $W_k^\top$ is implemented by forward warping.

We solve the Euler-Lagrange equation in (11) by a steepest descent (SD) solved by an explicit Euler scheme:

$$I_H^{n+1} = I_H^n + \tau\left(\sum_{k=1}^{N} W_k^T B D^T (I_L^{(k)} - DBW_k I_H^{(n)}) + \lambda \triangle I_H^n\right) \tag{12}$$

where τ is the time step. The two terms in the evolution of the high resolution image I_H induce a driving force that aims to match I_H (after warping, blurring and down-sampling) to all observations while imposing a linear diffusion of the intensities weighted by λ.

The whole algorithm including the accurate motion estimation (which is a very important aspect of super resolution) is summarized below.

Algorithm 1
Goal: Given a sequence of N - low resolution images $\{I_L^k\}_{k=1}^N$ estimate the in-

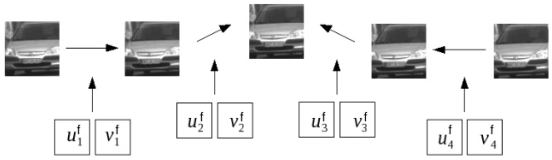


Fig. 4. Motion estimation between the frames

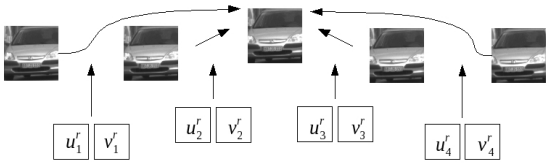


Fig. 5. Motion computation between the reference image and the frames.

terframe motion and infer a high resolution image I_H of the depicted scene.

1: Choose an image from the sequence as the reference image.
2: Estimate for each pair of consecutive frames the motion from one frame to the next (see Fig. 4) using the algorithm presented in Section 2 .
3: Using the motion fields u_i^f and v_i^f compute the motion fields u_i^r, v_i^r relative to the reference image (Fig. 5), the indices f (motion between mutual frames) and r (motion between reference frame and individual image) should indicate the difference of the motion maps.
4: Interpolate the motion fields u_i^r and v_i^r to the size of the image I_H.
5: Initialize I_H, by setting all pixel values to 0.
6: **for** t = 1 to T **do**
7: $sum := 0$;
8: **for** k = 1 to N **do**
9: $b := W_k I_H^t$ (backward warping);
10: $c := h(x, y) * b$ (convolution with the Gaussian kernel);
11: $c := Dc$ (down sampling to the size of I_L);
12: $d := (I_L^k - c)$;
13: $b := D^T d$ (up sampling without interpolation);
14: $c := h(x, y) * b$
15: $d := W_k^T c$ (forward warping);
16: $sum := sum + d$;
17: **end for**
18: $I_H^{t+1} = I_H^t + \tau(sum - \lambda \triangle I_H^t)$;
19: **end for**

The complete algorithm was implemented on the GPU. In Fig. 6 you can find the results, which were produced by *Algorithm 1.* As you can see there is a high spatial resolution improvement compared to the upscaling of a single frame. All characters on the licence plate are clearly identifiable. Nevertheless the resulting high resolution image is somewhat blurred, because L_2-regularizer does not allow

Fig. 6. A comparison between L^2-norm and one image from the sequence [21] upscaled by factor 2 shows obvious spatial resolution improvement. High resolution image is the result of *Algorithm 1* using a sequence of 10 input images. The parameters were set as $\lambda = 0.4$, time step $\tau = 0.01$ and iteration number $T = 150$

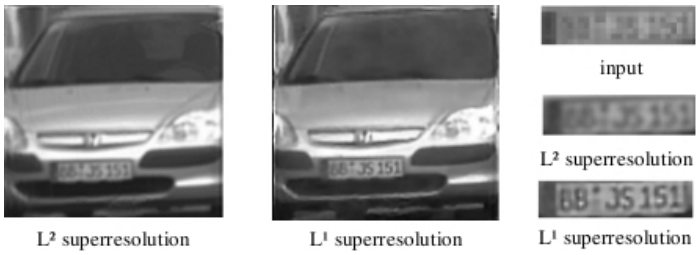


Fig. 7. A comparison between L^2-norm and L^1-norm shows that the L^1- norm allows to better preserve sharp edges in the super resolved image.

for discontinuities in high resolution image and it does not handle outliers that may arise from incorrectly estimated optical flow.

Robust Superresolution using L^1 Data and Regularity

In order to account for outliers and allow discontinuities in the reconstructed image, we replace data and regularity terms in (10) with respective L^1 expressions giving rise to the energy [20]:

$$E(I_H) = \sum_{k=1}^{N} \int_{\Omega} \left| DBW_k I_H - I_L^{(k)} \right| + \lambda \left| \nabla I_H \right| d\boldsymbol{x} \tag{13}$$

This gives rise to a gradient descent of the form:

$$\frac{\partial I_H}{\partial t} = \sum_{k=1}^{N} W_k^{\top} B^{\top} D^{\top} \frac{I_L^{(k)} - DBW_k I_H}{|I_L^{(k)} - DBW_k I_H|} + \lambda \operatorname{div}\left(\frac{\nabla I_H}{|\nabla I_H|} \right), \tag{14}$$

which is also implemented using an explicit Euler scheme – see equation (12). The robust regularizer gives rise to a nonlinear discontinuity preserving diffusion. For the numerical implementation we use a regularized differentiable approximation of the L^1 norm given by:

$$|s|_{\varepsilon} = \sqrt{|s|^2 + \varepsilon^2}.$$

The experimental results in (*Fig. 7-9*) demonstrate clearly that the L^1-formulation for motion-based super-resolution substantially improves the quality

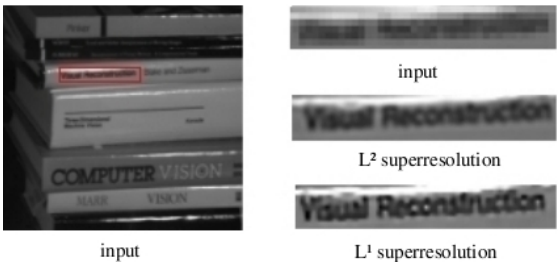


Fig. 8. While the L^2-norm allows a restoration of the image which is visibly better than the input images, the L^1-norm preserves sharp discontinuities even better. As input we used the book sequence from [15]

of the reconstructed image. Compared to the L^2-norm, we can see sharper edges. The numbers or letters that were not distinguishable in sequences are now clearly recognizable. The quality of the reconstructed super-resolution image depends on the accuracy of the estimated motion. In future research, we plan to investigate the joint optimization of intensity and motion field.

4 Conclusion

In this paper, we proposed a variational approach to super resolution which can handle arbitrary motion fields. In contrast to alternative super resolution approaches the motion field was not assumed as to be known. Instead we make use of a recently proposed dual decoupling scheme for high accuracy optic flow estimation. By minimizing a functional which depends on the input images and the estimated flow field we propose to invert the image formation process in order to compute a high resolution image of the filmed scene. We compared different variational approaches using L^2 and L^1 error norms for data and regularization term. This comparison shows that the L^1- norm is more robust to errors in motion and blur estimation and results in sharper super resolution images. Future work is focused on trying to simultaneously estimate the motion field and the super resolved image.

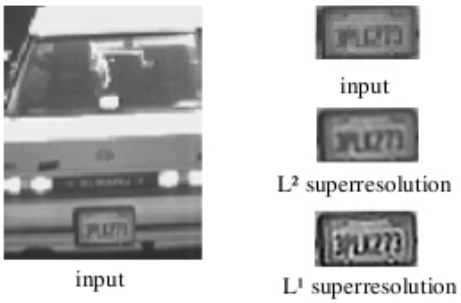


Fig. 9. Closeups show that the L^1-norm better preserves sharp edges in the restoration of the high resolution image. As input we used the car sequence from [14]

References

1. Chambolle, Antonin: An Algorithm for Total Variation Minimization and Applications. J. Math. Imaging Vis. **20** (2004) 89–97
2. Wedel A., Pock T., Zach C., Bischof H., Cremers, D.: An improved algorithm for TV-L1 optical flow computation. Proceedings of the Dagstuhl Visual Motion Analysis Workshop 2008.
3. Baker S., Scharstein D., Lewis, Roth, S., Black, M., Szeliski, R.: A Database and Evaluation Methodology for Optical Flow. http://vision.middlebury.edu/flow/data/
4. Elad, E., Feuer, A.: Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images. IEEE Trans. Image Processing **6** (1997) 1646–1658
5. Irani, M., Peleg, S.: Improving resolution by image registration. CVGIP: Graph. Models Image Process (1991) 231–239
6. Ur, H., Gross, D.: Improved resolution from subpixel shifted pictures. Graphical Models and Image Processing **54** (1992) 181–186
7. Farsiu, S., Robinson, M., Elad, M., Milanfar, P.: Fast and robust multiframe super resolution. IEEE Transactions on Image Processing **13** (2004) 1327–1344
8. Kim, S., Bose, N., Valenzuela, H.: Recursive reconstruction of high resolution image from noisy undersampled multiframes. IEEE Transactions on Acoustics, Speech and Signal Processing **38** (1990) 1013–1027
9. Huang T., Tsai R.: Multi-frame image restoration and registration Advances in Computer Vision and Image Processing **1** (1984) 317–339
10. Elad, M., Feuer, A.: Super-resolution reconstruction of image sequences IEEE Transactions on Pattern Analysis and Machine Intelligence **21** (1999) 817–834
11. Rudin, L., Osher, S., and Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60** (1992) 259–268
12. Pock, T.: Fast Total Variation for Computer Vision. Graz University of Technology, Austria (2008) PhD.
13. Zach, C., Pock, T., Bischof, H.: A Duality Based Approach for Realtime TV-L1 Optical Flow. Pattern Recognition Proc. DAGM Heidelberg, Germany (2007) 214–223
14. Malinfar, P: MDSP Super-Resolution And Demosaicing Datasets. University of California, Santa Cruz. http://www.ee.ucsc.edu/ milanfar/software/sr-datasets.html
15. Wang, C: Vision and Autonomous Systems Center's Image Database. Carnegie Mellon University. http://vasc.ri.cmu.edu/idb/html/motion/index.html
16. Protter, M., Elad, M., Takeda, H., Milanfar, P.: Generalizing the Non-Local-Means to Super-Resolution Reconstruction. IEEE Transactions on Image Processing, **18** (2009) 36–51
17. Ebrahimi, M., Vrscay, R.: Multi-Frame Super-Resolution with No Explicit Motion Estimation IPCV, (2008) 455–459
18. Kelley, C. T.: Iterative Methods for Linear and Nonlinear Equations. SIAM, Philadelphia, PA. (1995)
19. Zomet, A., Peleg, S.: Super-Resolution Imaging, chapter Superresolution from multiple images having arbitrary mutual motion Kluwer, (2001) 195–209.
20. Marquina, A., Osher, S. J.: Image Super-Resolution by TV-Regularization and Bregman Iteration. Journal of Scientific Computing, 37 **3**, (2008).
21. Daimler Research Stuttgart