Estimate, Compensate, Iterate: Joint Motion Estimation and Compensation in 4-D Cardiac C-arm Computed Tomography

Oliver Taubmann[1,2], Günter Lauritsch[3], Andreas Maier[1,2], Rebecca Fahrig[4], and Joachim Hornegger[1,2]

[1] Pattern Recognition Lab., Computer Science Department, Friedrich-Alexander-University Erlangen-Nuremberg, Germany
[2] Graduate School in Advanced Optical Technologies (SAOT), Erlangen, Germany
[3] Siemens AG, Healthcare, Forchheim, Germany
[4] Radiological Sciences Laboratory, Stanford University, California, USA

Abstract. C-arm computed tomography reconstruction of multiple cardiac phases could provide a highly useful tool to interventional cardiologists in the catheter laboratory. Today, however, for clinically reasonable acquisition protocols the achievable image quality is still severely limited due to undersampling artifacts. We propose an iterative optimization scheme combining image registration, motion compensation and spatio-temporal regularization to improve upon the state-of-the-art w. r. t. image quality and accuracy of motion estimation. Evaluation of clinical cases indicates an improved visual appearance and temporal consistency, evidenced by a strong decrease in temporal variance in uncontrasted regions accompanied by an increased sharpness of the contrasted left ventricular blood pool boundary. In a phantom study, the universal image quality index proposed by Wang et al. is raised from 0.80 to 0.95, with 1.0 corresponding to a perfect match with the ground truth. The results lay a promising foundation for interventional cardiac functional analysis.

1 Introduction

Ventricular wall motion analysis is commonly performed with cardiac magnetic resonance imaging or 3-D echocardiography. However, as these modalities do not readily fit into the workflow of most cardiac interventions, 2-D ventriculography is still frequently used for intraprocedural functional evaluation of cardiac motion. During cardiac resynchronization therapy, volumetric reconstruction of several heart phases, i. e. 4-D imaging, could help physicians find the optimal lead position by providing valuable information regarding the area of latest contraction [2]. Rotational angiography using C-arm devices as they are commonly found in catheter labs in combination with multi-segment retrospective ECG-gating could generate such images [4]. However, for protocols that achieve clinically reasonable acquisition times by performing only a single sweep [1], the gated reconstructions suffer severely from sparse view sampling.

N. Navab et al. (Eds.): MICCAI 2015, Part II, LNCS 9350, pp. 579–586, 2015.
DOI: 10.1007/978-3-319-24571-3_69

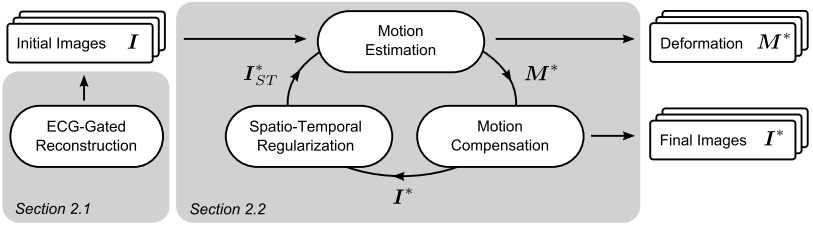

Fig. 1. A schematic overview of our iterative method, notation and paper structure.

In recent literature, several publications have focused on 4-D imaging with C-arm systems. For instance, Mory et al. employ 4-D iterative reconstruction with both spatial and temporal regularization [8]. While this approach can be of use in coping with angular undersampling, the achieved image quality is still severely limited by the small amount of data available for each cardiac phase and the "cartoon-like" appearance of the reconstructed images. Müller et al. attempt to overcome this problem by combining analytical reconstruction with motion compensation [10]. They estimate the motion between all phases of the initial, ECG-gated reconstruction, and then compensate for it in a final reconstruction step using all acquired projection data. However, motion estimation is highly sensitive to the quality of the initial images and artifacts typically propagate into the final images as artificial motion patterns. In this paper, we propose a method to combine the state-of-the-art motion-compensated filtered back-projection reconstruction technique with a novel iterative process incorporating spatio-temporal regularization to reduce artificial motion and further improve both image quality and accuracy of motion estimation.

2 Materials and Methods

Below, we first describe the generation of ECG-gated initial images (section 2.1). They serve as the input to our joint motion estimation and compensation framework presented in section 2.2. An overview of the method is shown in Fig. 1. Afterwards, details regarding our experimental data are provided (section 2.3).

2.1 Initial Image Generation

Initial images for each heart phase are generated using Feldkamp-Davis-Kress (FDK) filtered back-projection reconstruction combined with retrospective ECG-gating using a rectangular window. As only a subset of the projection images is used, we perform the following steps for denoising and reduction of few-view artifacts: **(i)** A two-pass removal of high-density objects such as catheters within the reconstruction field of view (FOV) [9]. **(ii)** A threshold-based masking in the filtered projection images to reduce artifacts caused by cables outside the FOV. **(iii)** An approach by McKinnon and Bates [7] to further reduce undersampling artifacts, in which object-dependent artifact images are estimated and

subsequently subtracted from the original images. **(iv)** Denoising with a joint bilateral filter [14] guided by a standard FDK reconstruction from all data.

2.2 Joint Motion Estimation and Compensation

Artifacts in the initial images propagate through motion estimation, leading to inconsistent movement patterns. Motion-compensated images, due to smoothness of the motion model and the no longer incomplete sampling, offer a clearer distinction of structures in the sense that high intensity differences are much more likely to be caused by an actual difference in underlying anatomical properties instead of merely artifacts. This enables us to use more aggressive edge-preserving filtering to eliminate inconsistency in both the spatial and temporal domain while preserving actual motion. In turn, these filtered images allow for a much more robust motion estimation, leading to a motion compensation with fewer inconsistencies. By iteratively filtering, estimating and compensating for motion, we can jointly improve both our images and our motion estimate.

Let $\boldsymbol{I} = \{\boldsymbol{I}^t : t \in \mathcal{N}_{\mathrm{ph}}\}$ denote 3-D images for all cardiac phases t, with $\mathcal{N}_{\mathrm{ph}} = \{1, \ldots, N_{\mathrm{phases}}\}$. Similarly, $\boldsymbol{M} = \{\boldsymbol{M}_{t\to t'} : t, t' \in \mathcal{N}_{\mathrm{ph}}\}$ are deformation operators, where $\boldsymbol{M}_{t\to t'}\boldsymbol{I}^t$ is the image $\boldsymbol{I}^t$ transformed to phase t', i. e. deformed according to the motion between cardiac phases t and t'. $\boldsymbol{M}_{t\to t}$ is defined as the identity transform and not part of the optimization process. The steps of motion estimation, compensation and spatio-temporal regularization described below are performed in an alternating manner.

Motion Estimation is performed by pair-wise 3-D/3-D image registration of all combinations of cardiac phases,

$$\boldsymbol{M}^* = \left\{ \underset{\boldsymbol{M}_{t\to t'}}{\operatorname{argmin}} \left[\mathcal{D}\left(\boldsymbol{I}^{t'}, \boldsymbol{M}_{t\to t'}\boldsymbol{I}^t \right) \right] : t, t' \in \mathcal{N}_{\mathrm{ph}},\ t \neq t' \right\}, \tag{1}$$

where $\mathcal{D}(\cdot,\cdot)$ is a dissimilarity measure. In our experiments, we use the negative normalized cross correlation and a uniform cubic B-spline motion model [10] with an isotropic control point spacing of 8 mm for the deformation operators $\boldsymbol{M}$. Optimization is performed on a multi-resolution pyramid with 3 levels using a quasi-Newton method, the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm. 10,000 random image samples per iteration and level are evaluated. For faster convergence, motion estimation is restricted to a manually defined region of interest (ROI) containing the heart and covering roughly 5 – 10 % of the full volume. Our implementation is based on `elastix`, a toolbox for nonrigid registration of medical images [3]. In the first iteration, $\boldsymbol{I}$ consists of the initial images (cf. section 2.1). In later iterations, we perform registration on the current regularized motion-compensated images, but do not initialize with the previous motion estimate so as not to introduce bias towards a solution potentially converged to an undesirable local minimum.

Motion Compensation amounts to incorporating the current motion estimate in the data fidelity equation of the reconstruction problem for all phases,

$$\boldsymbol{I}^* = \left\{ \boldsymbol{I}^t : \boldsymbol{A}^{t'}\boldsymbol{M}_{t\to t'}\boldsymbol{I}^t = \boldsymbol{P}^{t'};\ t, t' \in \mathcal{N}_{\mathrm{ph}} \right\}, \tag{2}$$

where $\boldsymbol{A}^{t'}$ is the X-ray projection operator belonging to the measured projection data $\boldsymbol{P}^{t'}$ assigned to phase t' during gating. Thus, for reconstruction of each individual phase t, data belonging to all phases t' are used. The images $\boldsymbol{I}^*$ are found using a variant of FDK-type filtered back-projection which compensates for motion by shifting each voxel according to the given transformation during the back-projection step [11]. This approach is efficient, which is crucial for use in an interventional setting, and at the same time guarantees that we do not lose resolution through repeated interpolation and regularization as we reconstruct anew—directly from the projection images—in each iteration.

Spatio-temporal Filtering is introduced to regularize the optimization. For this purpose, we adopt the common notion of favoring piece-wise constant images and apply spatial and temporal bilateral filters which are straightforward to parameterize and increase sparsity in the gradient domain. They read,

$$\boldsymbol{I}_S^t(\boldsymbol{x}) = \sum_{\boldsymbol{x'}\in N(\boldsymbol{x})} \frac{\boldsymbol{I}^t(\boldsymbol{x'})}{w_S} \cdot \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x'}\|_2^2}{2\sigma_S^2} - \frac{(\boldsymbol{I}^t(\boldsymbol{x})-\boldsymbol{I}^t(\boldsymbol{x'}))^2}{2\sigma_I^2}\right), \quad (3)$$

$$\boldsymbol{I}_{ST}^t(\boldsymbol{x}) = \sum_{t'\in\mathcal{N}_{\mathrm{ph}}} \frac{\boldsymbol{I}_S^{t'}(\boldsymbol{x})}{w_T} \cdot \exp\left(-\frac{\mathrm{dist}^2(t,t')}{2\sigma_T^2} - \frac{(\boldsymbol{I}^t(\boldsymbol{x})-\boldsymbol{I}^{t'}(\boldsymbol{x}))^2}{2\sigma_I^2}\right), \quad (4)$$

with σ_S, σ_T and σ_I the Gaussian standard deviations in the spatial, temporal and intensity domain, respectively. $N(\boldsymbol{x})$ is a local neighborhood around $\boldsymbol{x}$, w_S and w_T are normalization factors. $\mathrm{dist}(\cdot,\cdot)$ denotes the distance of two phases in the cardiac cycle, taking periodicity into account. By means of this edge-preserving filtering, we produce cleaner images $\boldsymbol{I}_{ST}^* = \left\{\boldsymbol{I}_{ST}^t : t \in \mathcal{N}_{\mathrm{ph}}\right\}$ with an improved temporal consistency for motion estimation in the next iteration. Note, however, that we stop iterating right after the motion compensation step to avoid an over-smoothed appearance in the final output.

2.3 Experiments

Clinical Data. Three data sets of clinical cases were acquired with an Artis zeego system (Siemens AG, Healthcare, Forchheim, Germany). 381 projection images were captured at approx. 30 Hz with an angular increment of 0.52° during one 14 s long rotation of the C-arm. The isotropic pixel resolution was 0.31 mm/pixel (0.21 in isocenter), the detector size 1240 × 960 pixels. External pacing was applied to ensure a heart rate of 115 bpm. The gating windows cover 10 % of the heart cycle each and are chosen to use all data without overlap, resulting in 10 cardiac phases to be reconstructed for both the initial and motion-compensated images. Undiluted contrast agent was injected in the pulmonary artery at a speed of 7 ml/s (91 ml total). All images were reconstructed on a grid of 256^3 voxels covering a volume of size $(25.6\,\mathrm{cm})^3$.

Phantom Model. For further evaluation, a numerical phantom data set was used [12,5]. 381 projection images of the phantom were simulated using a polychromatic spectrum, discretized in energy bins 2.5 keV wide from 25 keV to 120 keV (peak energy), and a time-current product of 2.5 mAs per X-ray pulse.

For bones and bone marrow, material properties match the mass attenuation coefficients found in the NIST X-ray table[1], while other structures were modeled with the absorption behavior of water for modified densities. For the contrasted left ventricular blood pool, the contrasted blood in the aorta and the myocardium, the densities were set to $2.5\,\mathrm{g/cm^3}$, $2.0\,\mathrm{g/cm^3}$ and $1.5\,\mathrm{g/cm^3}$, respectively. Additionaly, a complete set of projection images for a single cardiac phase (static phantom) was generated for reconstruction of a ground truth image. Other properties of the phantom data set, such as the number of heart beats and the resolution and dimensions of the images, were chosen to reflect the clinical data sets used in this study.

Experimental Setup. Images $\boldsymbol{I}_3^t(\boldsymbol{x})$ reconstructed with 3 iterations of the proposed method were compared to the the state-of-the-art method by Müller et al. [9], which corresponds to the result $\boldsymbol{I}_1^t(\boldsymbol{x})$ of the first motion-compensation step. As the preprocessing steps **(ii)** through **(iv)** (cf. section 2.1) have a strong impact on initial image quality, they were also added to their approach to enable a fair comparison with our method. In our experiments, we set $\sigma_S = 5\,\mathrm{mm}$ and $\sigma_T = 60\,\%$ of the cardiac cycle. Although rather large, these spatial and temporal kernel widths proved to work well for all our data sets as we choose σ_I so as to preserve the expected gray value difference between contrasted blood and surrounding tissue. In contrast to typical regularization weights, its value can be intuitively defined on the same scale as the observed intensities, which brings down the common, but often crucially unaddressed problem of proper parameterization to a manageable level.

Evaluation Measures. For quantitative evaluation of the phantom study, we compared the root mean square error (RMSE) as well as the universal image quality index (UQI), a correlation measure proposed by Wang et al. [15], w. r. t. the static ground truth reconstruction. For calculation of the UQI, the volume was divided into regular blocks of size $(16\,\mathrm{mm})^3$ each. Only those fully encompassed by the ROI used for registration (cf. section 2.2) were selected, resulting in a total of 9 blocks. To assess accuracy of motion estimation, we compared the dense motion fields obtained by evaluating the estimated B-spline model on the voxel grid against ground truth motion interpolated from the phantom surface definitions [6] between end-systole and end-diastole. For all data sets, overall visual image quality was scored by 9 experts in medical image processing at our institute. The images were shown in a randomized order to scorers blinded to the employed method. Similar to common approaches for (spatial) signal-to-noise ratio estimation, temporal consistency can also be measured as the temporal variance of intensity values in homogeneous regions consisting of uncontrasted blood and tissue: Ideally, they should appear mostly static over the course of the cardiac cycle as no motion is observed there. At the same time, it is essential to rule out a critical loss of spatial resolution when applying spatio-temporal smoothing. For nonlinear, object-dependent reconstruction methods, standardized modulation transfer function (MTF) evaluations cannot

[1] `http://physics.nist.gov/PhysRefData/Xcom/html/xcom1.html`

Table 1. For all data sets, the observed decrease in temporal variance, $\phi = 1 - \left(\mathrm{var}_t[\boldsymbol{I}_3^t]\right)\left(\mathrm{var}_t[\boldsymbol{I}_1^t]\right)^{-1}$, averaged over homogeneous uncontrasted, i. e. ideally static regions is given. Similarly, ξ_a and ξ_b are the ratios of edge sharpness estimates [13], i. e. descriptors of the relative sharpness increase, in short-axis slices towards the apical and basal ends of the end-diastolic heart, respectively. The visual image quality scores assigned by 9 experts at our institute are given as s_1 for $\boldsymbol{I}_1^t(\boldsymbol{x})$ and s_3 for $\boldsymbol{I}_3^t(\boldsymbol{x})$, ranging from 0 (unacceptable) to 4 (very good).

Data set	$\bar{\phi}$	ξ_a	ξ_b	s_1	s_3
Patient 1	50.1 %	1.22	1.24	2.1 ± 0.60	3.7 ± 0.50
Patient 2	31.5 %	1.02	1.21	1.4 ± 0.73	2.4 ± 0.88
Patient 3	30.8 %	1.03	1.22	1.4 ± 0.53	3.4 ± 0.73
Phantom	87.4 %	1.02	1.05	1.9 ± 0.78	3.7 ± 0.50

be applied, necessitating measurements of edge sharpness instead. Following the approach described in [13] for assessing reduction of motion blur, we measured the sharpness of the contrasted left ventricular blood pool as the anatomical structure of interest. For this purpose, ensembles of line profiles were placed along the edge in short-axis views and then used to robustly estimate its slope.

3 Results and Discussion

For the phantom data set, our method reduced the RMSE over the ROI from 86.3 HU to 73.6 HU (a decrease of 14.7 %), and raised the UQI from 0.80 to 0.95, where 1.0 would correspond to a perfect match. Although, like most, our phantom is piece-wise constant and therefore matches the prior knowledge assumption, rendering it easier to reconstruct with our method, we believe that it is still meaningful to compare the images as they are unregularized filtered backprojections of the same projection data albeit corrected according to different motion estimates. The motion estimation error, computed as the average magnitude of the difference vectors over the ROI, was reduced from 2.08 ± 1.92 mm to 1.52 ± 1.91 mm, a substantial decrease. Note that the numbers are higher than typically expected of target registration errors as we do not limit the evaluation to landmarks for which the motion is known to be observable in the images.

The results of the observer study, listed in Tab. 1, demonstrate a clear preference toward the proposed method as the images are disturbed less by artificial motion patterns. This is also reflected by the fact that our method was able to greatly reduce temporal inconsistency, as seen from the variance measurements in Tab. 1. The same effect is visualized by the difference images in Fig. 2 where we note a reduced amount of noise-like patterns. For visual comparison, animations of reconstructed data sets are available on our website[2]. From the edge sharpness measurements, which are also listed in Tab. 1, we observe a slight overall increase in sharpness. In the phantom, edges already exhibit a high sharpness with the

[2] `http://www5.cs.fau.de/our-team/taubmann-oliver/supplementary-material`

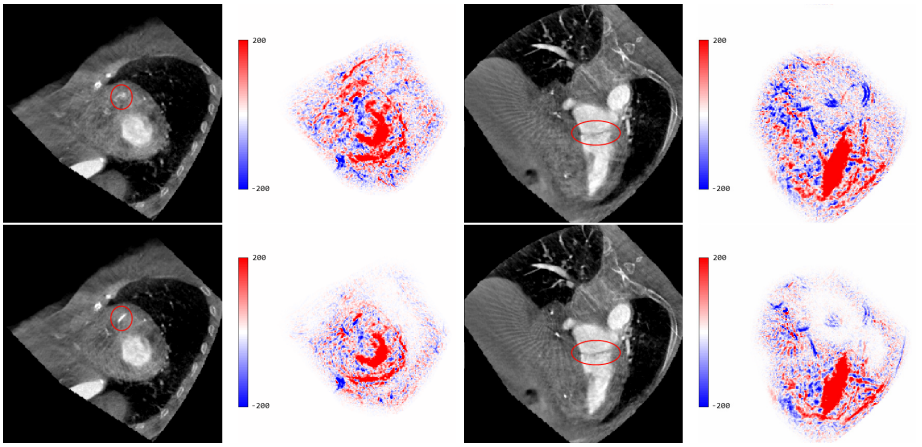

Fig. 2. Short-axis (left) and long-axis (right) views of patient data set 1 in end-diastole reconstructed with both the proposed (bottom) and the state-of-the-art method (top). The color-coded images show the difference between end-diastole and end-systole.

state-of-the-art method, leaving less room for improvement. The lowest values in the clinical data sets occur for the motion-intensive apical regions of patients 2 and 3, which we could confirm visually to have worse contrast than patient 1, thus requiring a more conservatively chosen value of σ_I. This also explains the lower decrease in temporal variance observed in these data sets.

Performance-wise, the motion estimation step constitutes the computational bottleneck of our current implementation. On average, registration of two volumes takes 10 - 30 s on a machine equipped with a quad-core 2.93 GHz CPU and 12 GB of RAM. Run times of the regularization and reconstruction steps are about 3 s and 85 s per volume, respectively.

4 Conclusion

For advancing accuracy and quality of interventional 4-D cardiac imaging, a scheme for iterative motion estimation and compensation is proposed, enabling the use of stronger regularization. Evaluations on C-arm CT data sets of clinical patients and a numerical heart phantom reveal distinct improvements compared to the state-of-the-art, both in terms of visual quality and accuracy of motion estimation. The results hold promise for efforts toward interventional cardiac functional analysis, which will also be subject of our future work.

Acknowledgments and Disclaimer. The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research Foundation (DFG) in the framework of the German excellence initiative, and thank Dr. Abt, Centre of Cardiovascular Diseases, Rotenburg a. d. Fulda, for providing the image data. The concepts and information presented in this paper are based on research and are not commercially available.

References

1. De Buck, S., Dauwe, D., Wielandts, J.Y., Claus, P., Janssens, S., Heidbuchel, H., Nuyens, D.: A new approach for prospectively gated cardiac rotational angiography. Proc. SPIE 8668 (2013)
2. Döring, M., Braunschweig, F., Eitel, C., Gaspar, T., Wetzel, U., Nitsche, B., Hindricks, G., Piorkowski, C.: Individually tailored left ventricular lead placement: lessons from multimodality integration between three-dimensional echocardiography and coronary sinus angiogram. Europace 15(5), 718–727 (2013)
3. Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J.: elastix: a toolbox for intensity based medical image registration. IEEE Trans. Med. Imag. 29(1), 196–205 (2010)
4. Lauritsch, G., Boese, J., Wigstrom, L., Kemeth, H., Fahrig, R.: Towards cardiac C-arm computed tomography. IEEE Trans. Med. Imag. 25(7), 922–934 (2006)
5. Maier, A., Hofmann, H., Berger, M., Fischer, P., Schwemmer, C., Wu, H., Müller, K., Hornegger, J., Choi, J.H., Riess, C., Keil, A., Fahrig, R.: CONRAD - A software framework for cone-beam imaging in radiology. Med. Phys. 40(11) (2013)
6. Maier, A., Taubmann, O., Wetzl, J., Wasza, J., Forman, C., Fischer, P., Hornegger, J., Fahrig, R.: Fast Interpolation of Dense Motion Fields from Synthetic Phantoms. In: BVM, pp. 168–173 (2014)
7. Mc Kinnon, G.C., Bates, R.: Towards imaging the beating heart usefully with a conventional CT scanner. IEEE Trans. Biomed. Eng. BME 28(2), 123–127 (1981)
8. Mory, C., Auvray, V., Zhang, B., Grass, M., Schäfer, D., Chen, S., Carroll, J., Rit, S., Peyrin, F., Douek, P., Boussel, L.: Cardiac C-arm computed tomography using a 3D + time ROI reconstruction method with spatial and temporal regularization. Med. Phys. 41, 021903 (2014)
9. Müller, K., Lauritsch, G., Schwemmer, C., Maier, A., Taubmann, O., Abt, B., Köhler, H., Nöttling, A., Hornegger, J., Fahrig, R.: Catheter artifact reduction (CAR) in dynamic cardiac chamber imaging with interventional C-arm CT. In: Proc. Intl. Conf. on Image Formation in X-ray CT, pp. 418–421 (2014)
10. Müller, K., Maier, A., Schwemmer, C., Lauritsch, G., Buck, S.D., Wielandts, J.Y., Hornegger, J., Fahrig, R.: Image artefact propagation in motion estimation and reconstruction in interventional cardiac C-arm CT. Phys. Med. Biol. 59(12), 3121–3138 (2014)
11. Schäfer, D., Borgert, J., Rasche, V., Grass, M.: Motion-compensated and gated cone beam filtered back-projection for 3-D rotational X-ray angiography. IEEE Trans. Med. Imag. 25, 898–906 (2006)
12. Segars, W.P., Sturgeon, G., Mendonca, S., Grimes, J., Tsui, B.M.W.: 4D XCAT. phantom for multimodality imaging research. Med. Phys. 37, 4902–4915 (2010)
13. Taubmann, O., Wetzl, J., Lauritsch, G., Maier, A., Hornegger, J.: Sharp as a Tack: Measuring and Comparing Edge Sharpness in Motion-Compensated Medical Image Reconstruction. In: BVM, pp. 425–430 (2015)
14. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: ICCV, pp. 839–846 (1998)
15. Wang, Z., Bovik, A.: A universal image quality index. IEEE Signal Proc. Let. 9(3), 81–84 (2002)